Part 1 – Exploratory Analysis

Taylor Gerkin

2024-06-15

Libraries were loaded (not included).

Next, data was imported and the summary reviewed.

train <- read.csv("train.csv")  
summary(train)

## id product\_code loading attribute\_0   
## Min. : 0 Length:26570 Min. : 33.16 Length:26570   
## 1st Qu.: 6642 Class :character 1st Qu.: 99.99 Class :character   
## Median :13284 Mode :character Median :122.39 Mode :character   
## Mean :13284 Mean :127.83   
## 3rd Qu.:19927 3rd Qu.:149.15   
## Max. :26569 Max. :385.86   
## NA's :250   
## attribute\_1 attribute\_2 attribute\_3 measurement\_0   
## Length:26570 Min. :5.000 Min. :5.00 Min. : 0.000   
## Class :character 1st Qu.:6.000 1st Qu.:6.00 1st Qu.: 4.000   
## Mode :character Median :6.000 Median :8.00 Median : 7.000   
## Mean :6.754 Mean :7.24 Mean : 7.416   
## 3rd Qu.:8.000 3rd Qu.:8.00 3rd Qu.:10.000   
## Max. :9.000 Max. :9.00 Max. :29.000   
##   
## measurement\_1 measurement\_2 measurement\_3 measurement\_4   
## Min. : 0.000 Min. : 0.000 Min. :13.97 Min. : 8.008   
## 1st Qu.: 5.000 1st Qu.: 4.000 1st Qu.:17.12 1st Qu.:11.051   
## Median : 8.000 Median : 6.000 Median :17.79 Median :11.733   
## Mean : 8.233 Mean : 6.257 Mean :17.79 Mean :11.732   
## 3rd Qu.:11.000 3rd Qu.: 8.000 3rd Qu.:18.47 3rd Qu.:12.410   
## Max. :29.000 Max. :24.000 Max. :21.50 Max. :16.484   
## NA's :381 NA's :538   
## measurement\_5 measurement\_6 measurement\_7 measurement\_8   
## Min. :12.07 Min. :12.71 Min. : 7.968 Min. :15.22   
## 1st Qu.:16.44 1st Qu.:16.84 1st Qu.:11.045 1st Qu.:18.34   
## Median :17.13 Median :17.52 Median :11.712 Median :19.02   
## Mean :17.13 Mean :17.51 Mean :11.717 Mean :19.02   
## 3rd Qu.:17.80 3rd Qu.:18.18 3rd Qu.:12.391 3rd Qu.:19.71   
## Max. :21.43 Max. :21.54 Max. :15.419 Max. :23.81   
## NA's :676 NA's :796 NA's :937 NA's :1048   
## measurement\_9 measurement\_10 measurement\_11 measurement\_12   
## Min. : 7.537 Min. : 9.323 Min. :12.46 Min. : 5.167   
## 1st Qu.:10.757 1st Qu.:15.209 1st Qu.:18.17 1st Qu.:10.703   
## Median :11.430 Median :16.127 Median :19.21 Median :11.717   
## Mean :11.431 Mean :16.118 Mean :19.17 Mean :11.703   
## 3rd Qu.:12.102 3rd Qu.:17.025 3rd Qu.:20.21 3rd Qu.:12.709   
## Max. :15.412 Max. :22.479 Max. :25.64 Max. :17.663   
## NA's :1227 NA's :1300 NA's :1468 NA's :1601   
## measurement\_13 measurement\_14 measurement\_15 measurement\_16   
## Min. :10.89 Min. : 9.14 Min. : 9.104 Min. : 9.701   
## 1st Qu.:14.89 1st Qu.:15.06 1st Qu.:13.957 1st Qu.:15.268   
## Median :15.63 Median :16.04 Median :14.969 Median :16.436   
## Mean :15.65 Mean :16.05 Mean :14.996 Mean :16.461   
## 3rd Qu.:16.37 3rd Qu.:17.08 3rd Qu.:16.018 3rd Qu.:17.628   
## Max. :22.71 Max. :22.30 Max. :21.626 Max. :24.094   
## NA's :1774 NA's :1874 NA's :2009 NA's :2110   
## measurement\_17 failure   
## Min. : 196.8 Length:26570   
## 1st Qu.: 619.0 Class :character   
## Median : 701.0 Mode :character   
## Mean : 701.3   
## 3rd Qu.: 784.1   
## Max. :1312.8   
## NA's :2284

str(train)

## 'data.frame': 26570 obs. of 26 variables:  
## $ id : int 0 1 2 3 4 5 6 7 8 9 ...  
## $ product\_code : chr "A" "A" "A" "A" ...  
## $ loading : num 80.1 84.9 82.4 101.1 188.1 ...  
## $ attribute\_0 : chr "material\_7" "material\_7" "material\_7" "material\_7" ...  
## $ attribute\_1 : chr "material\_8" "material\_8" "material\_8" "material\_8" ...  
## $ attribute\_2 : int 9 9 9 9 9 9 9 9 9 9 ...  
## $ attribute\_3 : int 5 5 5 5 5 5 5 5 5 5 ...  
## $ measurement\_0 : int 7 14 12 13 9 11 12 4 9 10 ...  
## $ measurement\_1 : int 8 3 1 2 2 4 2 8 6 4 ...  
## $ measurement\_2 : int 4 3 5 6 8 0 4 8 5 7 ...  
## $ measurement\_3 : num 18 18.2 18.1 17.3 19.3 ...  
## $ measurement\_4 : num 12.5 11.5 11.7 11.2 12.9 ...  
## $ measurement\_5 : num 15.7 17.7 16.7 18.6 17 ...  
## $ measurement\_6 : num 19.3 17.9 18.2 18.3 15.7 ...  
## $ measurement\_7 : num 11.7 12.7 12.7 12.6 11.3 ...  
## $ measurement\_8 : num 20.2 17.9 18.3 19.1 18.1 ...  
## $ measurement\_9 : num 10.7 12.4 12.7 12.5 10.3 ...  
## $ measurement\_10: num 15.9 17.9 15.6 16.3 17.1 ...  
## $ measurement\_11: num 17.6 17.9 NA 18.4 19.9 ...  
## $ measurement\_12: num 15.2 11.8 13.8 10 12.4 ...  
## $ measurement\_13: num 15 14.7 16.7 15.2 16.2 ...  
## $ measurement\_14: num NA 15.4 18.6 15.6 12.8 ...  
## $ measurement\_15: num 13 14.4 14.1 16.2 13.2 ...  
## $ measurement\_16: num 14.7 15.6 17.9 17.2 16.4 ...  
## $ measurement\_17: num 764 682 663 826 580 ...  
## $ failure : chr "No" "No" "No" "No" ...

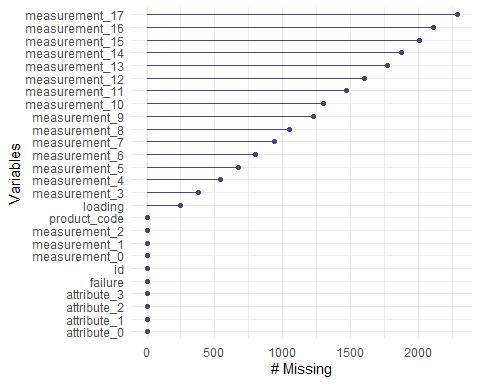
Character variables were transformed to factors.

train = train %>% mutate(failure = as\_factor(failure)) %>% mutate(product\_code = as\_factor(product\_code)) %>% mutate(attribute\_0 = as\_factor(attribute\_0)) %>% mutate(attribute\_1 = as\_factor(attribute\_1))  
  
summary(train)

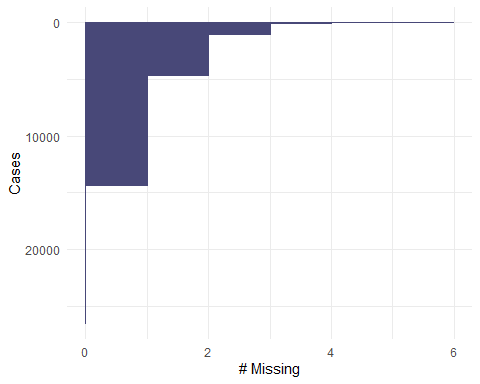
## id product\_code loading attribute\_0   
## Min. : 0 A:5100 Min. : 33.16 material\_7:21320   
## 1st Qu.: 6642 B:5250 1st Qu.: 99.99 material\_5: 5250   
## Median :13284 C:5765 Median :122.39   
## Mean :13284 D:5112 Mean :127.83   
## 3rd Qu.:19927 E:5343 3rd Qu.:149.15   
## Max. :26569 Max. :385.86   
## NA's :250   
## attribute\_1 attribute\_2 attribute\_3 measurement\_0   
## material\_8:10865 Min. :5.000 Min. :5.00 Min. : 0.000   
## material\_5:10362 1st Qu.:6.000 1st Qu.:6.00 1st Qu.: 4.000   
## material\_6: 5343 Median :6.000 Median :8.00 Median : 7.000   
## Mean :6.754 Mean :7.24 Mean : 7.416   
## 3rd Qu.:8.000 3rd Qu.:8.00 3rd Qu.:10.000   
## Max. :9.000 Max. :9.00 Max. :29.000   
##   
## measurement\_1 measurement\_2 measurement\_3 measurement\_4   
## Min. : 0.000 Min. : 0.000 Min. :13.97 Min. : 8.008   
## 1st Qu.: 5.000 1st Qu.: 4.000 1st Qu.:17.12 1st Qu.:11.051   
## Median : 8.000 Median : 6.000 Median :17.79 Median :11.733   
## Mean : 8.233 Mean : 6.257 Mean :17.79 Mean :11.732   
## 3rd Qu.:11.000 3rd Qu.: 8.000 3rd Qu.:18.47 3rd Qu.:12.410   
## Max. :29.000 Max. :24.000 Max. :21.50 Max. :16.484   
## NA's :381 NA's :538   
## measurement\_5 measurement\_6 measurement\_7 measurement\_8   
## Min. :12.07 Min. :12.71 Min. : 7.968 Min. :15.22   
## 1st Qu.:16.44 1st Qu.:16.84 1st Qu.:11.045 1st Qu.:18.34   
## Median :17.13 Median :17.52 Median :11.712 Median :19.02   
## Mean :17.13 Mean :17.51 Mean :11.717 Mean :19.02   
## 3rd Qu.:17.80 3rd Qu.:18.18 3rd Qu.:12.391 3rd Qu.:19.71   
## Max. :21.43 Max. :21.54 Max. :15.419 Max. :23.81   
## NA's :676 NA's :796 NA's :937 NA's :1048   
## measurement\_9 measurement\_10 measurement\_11 measurement\_12   
## Min. : 7.537 Min. : 9.323 Min. :12.46 Min. : 5.167   
## 1st Qu.:10.757 1st Qu.:15.209 1st Qu.:18.17 1st Qu.:10.703   
## Median :11.430 Median :16.127 Median :19.21 Median :11.717   
## Mean :11.431 Mean :16.118 Mean :19.17 Mean :11.703   
## 3rd Qu.:12.102 3rd Qu.:17.025 3rd Qu.:20.21 3rd Qu.:12.709   
## Max. :15.412 Max. :22.479 Max. :25.64 Max. :17.663   
## NA's :1227 NA's :1300 NA's :1468 NA's :1601   
## measurement\_13 measurement\_14 measurement\_15 measurement\_16   
## Min. :10.89 Min. : 9.14 Min. : 9.104 Min. : 9.701   
## 1st Qu.:14.89 1st Qu.:15.06 1st Qu.:13.957 1st Qu.:15.268   
## Median :15.63 Median :16.04 Median :14.969 Median :16.436   
## Mean :15.65 Mean :16.05 Mean :14.996 Mean :16.461   
## 3rd Qu.:16.37 3rd Qu.:17.08 3rd Qu.:16.018 3rd Qu.:17.628   
## Max. :22.71 Max. :22.30 Max. :21.626 Max. :24.094   
## NA's :1774 NA's :1874 NA's :2009 NA's :2110   
## measurement\_17 failure   
## Min. : 196.8 No :20921   
## 1st Qu.: 619.0 Yes: 5649   
## Median : 701.0   
## Mean : 701.3   
## 3rd Qu.: 784.1   
## Max. :1312.8   
## NA's :2284

Began to visualize missing data.

gg\_miss\_var(train)

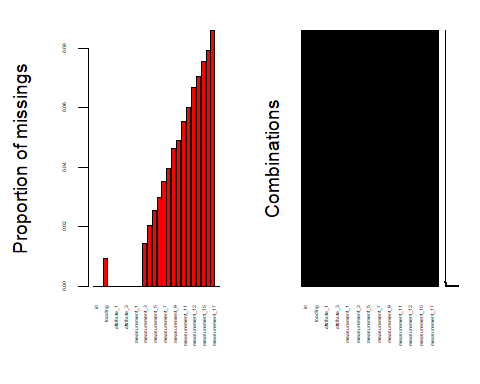


gg\_miss\_case(train)

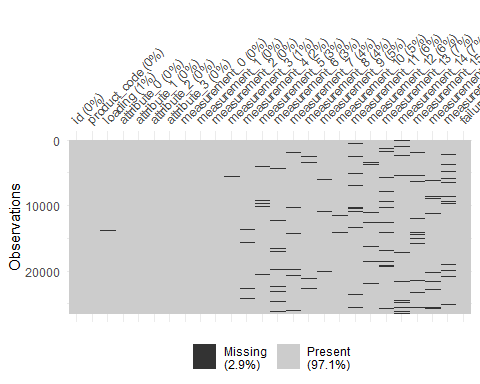


vim\_plot = aggr(train, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.3)

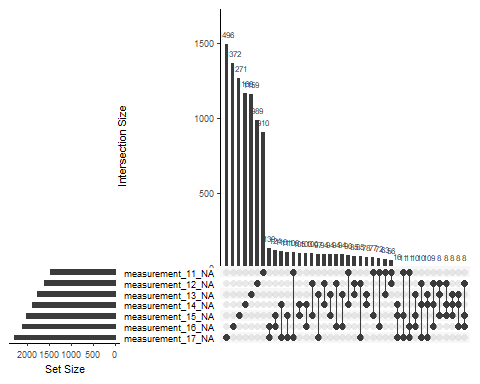
## Warning in plot.aggr(res, ...): not enough vertical space to display  
## frequencies (too many combinations)



vis\_miss(train)



gg\_miss\_upset(train, nsets = 7)



set.seed(1234)   
imp\_failure = mice(train, m=5, method='pmm', printFlag=FALSE)

## Warning: Number of logged events: 756

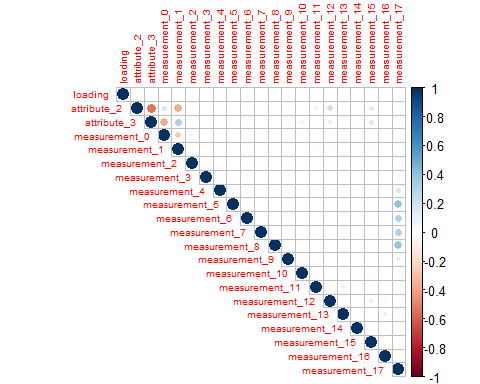
summary(imp\_failure)

## Class: mids  
## Number of multiple imputations: 5   
## Imputation methods:  
## id product\_code loading attribute\_0 attribute\_1   
## "" "" "pmm" "" ""   
## attribute\_2 attribute\_3 measurement\_0 measurement\_1 measurement\_2   
## "" "" "" "" ""   
## measurement\_3 measurement\_4 measurement\_5 measurement\_6 measurement\_7   
## "pmm" "pmm" "pmm" "pmm" "pmm"   
## measurement\_8 measurement\_9 measurement\_10 measurement\_11 measurement\_12   
## "pmm" "pmm" "pmm" "pmm" "pmm"   
## measurement\_13 measurement\_14 measurement\_15 measurement\_16 measurement\_17   
## "pmm" "pmm" "pmm" "pmm" "pmm"   
## failure   
## ""   
## PredictorMatrix:  
## id product\_code loading attribute\_0 attribute\_1 attribute\_2  
## id 0 1 1 1 1 1  
## product\_code 1 0 1 1 1 1  
## loading 1 1 0 1 1 1  
## attribute\_0 1 1 1 0 1 1  
## attribute\_1 1 1 1 1 0 1  
## attribute\_2 1 1 1 1 1 0  
## attribute\_3 measurement\_0 measurement\_1 measurement\_2  
## id 1 1 1 1  
## product\_code 1 1 1 1  
## loading 1 1 1 1  
## attribute\_0 1 1 1 1  
## attribute\_1 1 1 1 1  
## attribute\_2 1 1 1 1  
## measurement\_3 measurement\_4 measurement\_5 measurement\_6  
## id 1 1 1 1  
## product\_code 1 1 1 1  
## loading 1 1 1 1  
## attribute\_0 1 1 1 1  
## attribute\_1 1 1 1 1  
## attribute\_2 1 1 1 1  
## measurement\_7 measurement\_8 measurement\_9 measurement\_10  
## id 1 1 1 1  
## product\_code 1 1 1 1  
## loading 1 1 1 1  
## attribute\_0 1 1 1 1  
## attribute\_1 1 1 1 1  
## attribute\_2 1 1 1 1  
## measurement\_11 measurement\_12 measurement\_13 measurement\_14  
## id 1 1 1 1  
## product\_code 1 1 1 1  
## loading 1 1 1 1  
## attribute\_0 1 1 1 1  
## attribute\_1 1 1 1 1  
## attribute\_2 1 1 1 1  
## measurement\_15 measurement\_16 measurement\_17 failure  
## id 1 1 1 1  
## product\_code 1 1 1 1  
## loading 1 1 1 1  
## attribute\_0 1 1 1 1  
## attribute\_1 1 1 1 1  
## attribute\_2 1 1 1 1  
## Number of logged events: 756   
## it im dep meth  
## 1 1 1 loading pmm  
## 2 1 1 loading pmm  
## 3 1 1 measurement\_3 pmm  
## 4 1 1 measurement\_3 pmm  
## 5 1 1 measurement\_4 pmm  
## 6 1 1 measurement\_4 pmm  
## out  
## 1 product\_codeB, product\_codeD, attribute\_0material\_5, attribute\_1material\_5, attribute\_3  
## 2 mice detected that your data are (nearly) multi-collinear.\nIt applied a ridge penalty to continue calculations, but the results can be unstable.\nDoes your dataset contain duplicates, linear transformation, or factors with unique respondent names?  
## 3 product\_codeB, loading, attribute\_1material\_6, measurement\_0, measurement\_4  
## 4 mice detected that your data are (nearly) multi-collinear.\nIt applied a ridge penalty to continue calculations, but the results can be unstable.\nDoes your dataset contain duplicates, linear transformation, or factors with unique respondent names?  
## 5 product\_codeB, attribute\_1material\_5, attribute\_1material\_6, measurement\_0, measurement\_2  
## 6 mice detected that your data are (nearly) multi-collinear.\nIt applied a ridge penalty to continue calculations, but the results can be unstable.\nDoes your dataset contain duplicates, linear transformation, or factors with unique respondent names?

train\_complete = complete(imp\_failure)   
summary(train\_complete)

## id product\_code loading attribute\_0   
## Min. : 0 A:5100 Min. : 33.16 material\_7:21320   
## 1st Qu.: 6642 B:5250 1st Qu.: 99.96 material\_5: 5250   
## Median :13284 C:5765 Median :122.36   
## Mean :13284 D:5112 Mean :127.82   
## 3rd Qu.:19927 E:5343 3rd Qu.:149.16   
## Max. :26569 Max. :385.86   
## attribute\_1 attribute\_2 attribute\_3 measurement\_0   
## material\_8:10865 Min. :5.000 Min. :5.00 Min. : 0.000   
## material\_5:10362 1st Qu.:6.000 1st Qu.:6.00 1st Qu.: 4.000   
## material\_6: 5343 Median :6.000 Median :8.00 Median : 7.000   
## Mean :6.754 Mean :7.24 Mean : 7.416   
## 3rd Qu.:8.000 3rd Qu.:8.00 3rd Qu.:10.000   
## Max. :9.000 Max. :9.00 Max. :29.000   
## measurement\_1 measurement\_2 measurement\_3 measurement\_4   
## Min. : 0.000 Min. : 0.000 Min. :13.97 Min. : 8.008   
## 1st Qu.: 5.000 1st Qu.: 4.000 1st Qu.:17.12 1st Qu.:11.013   
## Median : 8.000 Median : 6.000 Median :17.79 Median :11.716   
## Mean : 8.233 Mean : 6.257 Mean :17.80 Mean :11.710   
## 3rd Qu.:11.000 3rd Qu.: 8.000 3rd Qu.:18.48 3rd Qu.:12.405   
## Max. :29.000 Max. :24.000 Max. :21.50 Max. :16.484   
## measurement\_5 measurement\_6 measurement\_7 measurement\_8   
## Min. :12.07 Min. :12.71 Min. : 7.968 Min. :15.22   
## 1st Qu.:16.39 1st Qu.:16.84 1st Qu.:11.049 1st Qu.:18.38   
## Median :17.10 Median :17.52 Median :11.713 Median :19.07   
## Mean :17.08 Mean :17.51 Mean :11.717 Mean :19.12   
## 3rd Qu.:17.79 3rd Qu.:18.18 3rd Qu.:12.391 3rd Qu.:19.81   
## Max. :21.43 Max. :21.54 Max. :15.419 Max. :23.81   
## measurement\_9 measurement\_10 measurement\_11 measurement\_12   
## Min. : 7.537 Min. : 9.323 Min. :12.46 Min. : 5.167   
## 1st Qu.:10.734 1st Qu.:15.173 1st Qu.:18.24 1st Qu.:10.715   
## Median :11.420 Median :16.110 Median :19.32 Median :11.690   
## Mean :11.414 Mean :16.083 Mean :19.28 Mean :11.695   
## 3rd Qu.:12.065 3rd Qu.:16.975 3rd Qu.:20.35 3rd Qu.:12.699   
## Max. :15.412 Max. :22.479 Max. :25.64 Max. :17.663   
## measurement\_13 measurement\_14 measurement\_15 measurement\_16   
## Min. :10.89 Min. : 9.14 Min. : 9.104 Min. : 9.701   
## 1st Qu.:14.85 1st Qu.:15.08 1st Qu.:13.820 1st Qu.:15.263   
## Median :15.65 Median :16.07 Median :14.806 Median :16.534   
## Mean :15.64 Mean :16.09 Mean :14.812 Mean :16.718   
## 3rd Qu.:16.31 3rd Qu.:17.13 3rd Qu.:15.918 3rd Qu.:17.846   
## Max. :22.71 Max. :22.30 Max. :21.626 Max. :24.094   
## measurement\_17 failure   
## Min. : 196.8 No :20921   
## 1st Qu.: 614.1 Yes: 5649   
## Median : 698.6   
## Mean : 697.3   
## 3rd Qu.: 783.6   
## Max. :1312.8

train\_complete\_numeric <- train\_complete %>% select\_if(is.numeric)  
train\_complete\_numeric <- train\_complete\_numeric %>% select(-id)  
train\_complete\_numeric.cor = cor(train\_complete\_numeric)  
corrplot(train\_complete\_numeric.cor, type = 'upper', tl.cex = .6)



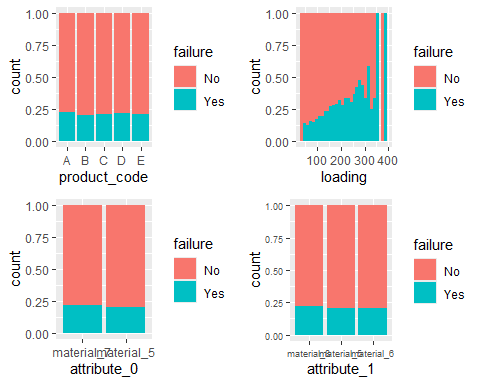
rcorr(as.matrix(train\_complete\_numeric))

## loading attribute\_2 attribute\_3 measurement\_0 measurement\_1  
## loading 1.00 -0.01 0.00 -0.01 0.00  
## attribute\_2 -0.01 1.00 -0.54 0.18 -0.38  
## attribute\_3 0.00 -0.54 1.00 -0.37 0.33  
## measurement\_0 -0.01 0.18 -0.37 1.00 -0.26  
## measurement\_1 0.00 -0.38 0.33 -0.26 1.00  
## measurement\_2 0.00 0.03 0.01 -0.08 0.01  
## measurement\_3 -0.01 0.00 0.00 0.00 -0.01  
## measurement\_4 -0.01 -0.01 0.01 -0.01 0.00  
## measurement\_5 -0.01 0.00 0.01 0.00 0.00  
## measurement\_6 0.01 0.00 0.00 0.00 0.00  
## measurement\_7 0.00 0.01 -0.01 0.00 -0.02  
## measurement\_8 0.00 0.00 0.00 -0.01 0.00  
## measurement\_9 -0.01 0.00 0.00 0.01 0.01  
## measurement\_10 -0.01 -0.01 -0.10 0.03 0.01  
## measurement\_11 0.00 0.09 -0.03 0.00 -0.04  
## measurement\_12 0.00 0.21 -0.12 0.03 -0.08  
## measurement\_13 0.01 0.05 0.03 -0.03 0.01  
## measurement\_14 0.00 0.02 -0.04 0.02 -0.01  
## measurement\_15 -0.01 -0.14 0.16 -0.05 0.05  
## measurement\_16 -0.01 0.03 -0.03 0.00 0.00  
## measurement\_17 0.01 -0.03 -0.03 0.02 -0.01  
## measurement\_2 measurement\_3 measurement\_4 measurement\_5  
## loading 0.00 -0.01 -0.01 -0.01  
## attribute\_2 0.03 0.00 -0.01 0.00  
## attribute\_3 0.01 0.00 0.01 0.01  
## measurement\_0 -0.08 0.00 -0.01 0.00  
## measurement\_1 0.01 -0.01 0.00 0.00  
## measurement\_2 1.00 0.00 -0.01 0.00  
## measurement\_3 0.00 1.00 0.01 0.01  
## measurement\_4 -0.01 0.01 1.00 0.00  
## measurement\_5 0.00 0.01 0.00 1.00  
## measurement\_6 -0.01 0.01 0.00 0.01  
## measurement\_7 0.00 0.01 0.00 -0.01  
## measurement\_8 0.00 0.00 0.01 0.00  
## measurement\_9 0.00 0.01 0.01 0.00  
## measurement\_10 0.02 0.00 -0.01 0.00  
## measurement\_11 0.00 -0.01 0.00 0.01  
## measurement\_12 0.00 0.00 0.00 0.00  
## measurement\_13 0.03 0.01 0.00 -0.01  
## measurement\_14 0.01 0.00 0.00 0.00  
## measurement\_15 0.00 -0.01 0.00 0.00  
## measurement\_16 0.02 0.00 0.00 -0.01  
## measurement\_17 -0.02 0.07 0.17 0.39  
## measurement\_6 measurement\_7 measurement\_8 measurement\_9  
## loading 0.01 0.00 0.00 -0.01  
## attribute\_2 0.00 0.01 0.00 0.00  
## attribute\_3 0.00 -0.01 0.00 0.00  
## measurement\_0 0.00 0.00 -0.01 0.01  
## measurement\_1 0.00 -0.02 0.00 0.01  
## measurement\_2 -0.01 0.00 0.00 0.00  
## measurement\_3 0.01 0.01 0.00 0.01  
## measurement\_4 0.00 0.00 0.01 0.01  
## measurement\_5 0.01 -0.01 0.00 0.00  
## measurement\_6 1.00 0.01 0.00 0.01  
## measurement\_7 0.01 1.00 0.00 0.00  
## measurement\_8 0.00 0.00 1.00 -0.01  
## measurement\_9 0.01 0.00 -0.01 1.00  
## measurement\_10 0.00 0.00 0.00 0.00  
## measurement\_11 0.00 -0.01 -0.01 -0.01  
## measurement\_12 -0.01 0.00 0.00 -0.01  
## measurement\_13 0.00 0.00 0.00 -0.01  
## measurement\_14 0.00 0.00 0.00 -0.01  
## measurement\_15 0.00 0.00 0.01 0.00  
## measurement\_16 0.00 -0.01 0.01 0.00  
## measurement\_17 0.30 0.30 0.39 0.13  
## measurement\_10 measurement\_11 measurement\_12 measurement\_13  
## loading -0.01 0.00 0.00 0.01  
## attribute\_2 -0.01 0.09 0.21 0.05  
## attribute\_3 -0.10 -0.03 -0.12 0.03  
## measurement\_0 0.03 0.00 0.03 -0.03  
## measurement\_1 0.01 -0.04 -0.08 0.01  
## measurement\_2 0.02 0.00 0.00 0.03  
## measurement\_3 0.00 -0.01 0.00 0.01  
## measurement\_4 -0.01 0.00 0.00 0.00  
## measurement\_5 0.00 0.01 0.00 -0.01  
## measurement\_6 0.00 0.00 -0.01 0.00  
## measurement\_7 0.00 -0.01 0.00 0.00  
## measurement\_8 0.00 -0.01 0.00 0.00  
## measurement\_9 0.00 -0.01 -0.01 -0.01  
## measurement\_10 1.00 -0.03 0.03 0.03  
## measurement\_11 -0.03 1.00 0.08 -0.08  
## measurement\_12 0.03 0.08 1.00 0.00  
## measurement\_13 0.03 -0.08 0.00 1.00  
## measurement\_14 0.04 -0.07 -0.02 0.00  
## measurement\_15 -0.04 -0.03 -0.12 0.02  
## measurement\_16 0.00 0.06 0.07 0.10  
## measurement\_17 0.00 -0.01 -0.01 -0.02  
## measurement\_14 measurement\_15 measurement\_16 measurement\_17  
## loading 0.00 -0.01 -0.01 0.01  
## attribute\_2 0.02 -0.14 0.03 -0.03  
## attribute\_3 -0.04 0.16 -0.03 -0.03  
## measurement\_0 0.02 -0.05 0.00 0.02  
## measurement\_1 -0.01 0.05 0.00 -0.01  
## measurement\_2 0.01 0.00 0.02 -0.02  
## measurement\_3 0.00 -0.01 0.00 0.07  
## measurement\_4 0.00 0.00 0.00 0.17  
## measurement\_5 0.00 0.00 -0.01 0.39  
## measurement\_6 0.00 0.00 0.00 0.30  
## measurement\_7 0.00 0.00 -0.01 0.30  
## measurement\_8 0.00 0.01 0.01 0.39  
## measurement\_9 -0.01 0.00 0.00 0.13  
## measurement\_10 0.04 -0.04 0.00 0.00  
## measurement\_11 -0.07 -0.03 0.06 -0.01  
## measurement\_12 -0.02 -0.12 0.07 -0.01  
## measurement\_13 0.00 0.02 0.10 -0.02  
## measurement\_14 1.00 0.02 -0.06 0.00  
## measurement\_15 0.02 1.00 0.01 0.00  
## measurement\_16 -0.06 0.01 1.00 -0.01  
## measurement\_17 0.00 0.00 -0.01 1.00  
##   
## n= 26570   
##   
##   
## P  
## loading attribute\_2 attribute\_3 measurement\_0 measurement\_1  
## loading 0.2164 0.9061 0.3428 0.7332   
## attribute\_2 0.2164 0.0000 0.0000 0.0000   
## attribute\_3 0.9061 0.0000 0.0000 0.0000   
## measurement\_0 0.3428 0.0000 0.0000 0.0000   
## measurement\_1 0.7332 0.0000 0.0000 0.0000   
## measurement\_2 0.9971 0.0000 0.0307 0.0000 0.4121   
## measurement\_3 0.3447 0.6234 0.6768 0.4756 0.3427   
## measurement\_4 0.0577 0.1124 0.0167 0.2827 0.9063   
## measurement\_5 0.2443 0.9289 0.1801 0.6198 0.8808   
## measurement\_6 0.3261 0.8921 0.9896 0.5703 0.9024   
## measurement\_7 0.5952 0.0776 0.1612 0.7408 0.0049   
## measurement\_8 0.9172 0.6543 0.6937 0.2169 0.8668   
## measurement\_9 0.2784 0.9381 0.7808 0.2048 0.4104   
## measurement\_10 0.2416 0.4044 0.0000 0.0000 0.0709   
## measurement\_11 0.7794 0.0000 0.0000 0.6921 0.0000   
## measurement\_12 0.8625 0.0000 0.0000 0.0000 0.0000   
## measurement\_13 0.2841 0.0000 0.0000 0.0000 0.2525   
## measurement\_14 0.6975 0.0123 0.0000 0.0021 0.1050   
## measurement\_15 0.3597 0.0000 0.0000 0.0000 0.0000   
## measurement\_16 0.0538 0.0000 0.0000 0.5397 0.4497   
## measurement\_17 0.0314 0.0000 0.0000 0.0006 0.1922   
## measurement\_2 measurement\_3 measurement\_4 measurement\_5  
## loading 0.9971 0.3447 0.0577 0.2443   
## attribute\_2 0.0000 0.6234 0.1124 0.9289   
## attribute\_3 0.0307 0.6768 0.0167 0.1801   
## measurement\_0 0.0000 0.4756 0.2827 0.6198   
## measurement\_1 0.4121 0.3427 0.9063 0.8808   
## measurement\_2 0.6780 0.2230 0.8679   
## measurement\_3 0.6780 0.0918 0.1704   
## measurement\_4 0.2230 0.0918 0.4431   
## measurement\_5 0.8679 0.1704 0.4431   
## measurement\_6 0.1885 0.3255 0.7115 0.3329   
## measurement\_7 0.6511 0.3228 0.5452 0.2179   
## measurement\_8 0.5388 0.5652 0.0433 0.7561   
## measurement\_9 0.8064 0.2155 0.0935 0.7520   
## measurement\_10 0.0037 0.9619 0.0364 0.7463   
## measurement\_11 0.7566 0.2761 0.9728 0.0562   
## measurement\_12 0.9376 0.5404 0.6448 0.6545   
## measurement\_13 0.0000 0.3393 0.8922 0.0420   
## measurement\_14 0.0506 0.6333 0.9011 0.5417   
## measurement\_15 0.9660 0.1436 0.4171 0.7452   
## measurement\_16 0.0002 0.9992 0.4527 0.1927   
## measurement\_17 0.0052 0.0000 0.0000 0.0000   
## measurement\_6 measurement\_7 measurement\_8 measurement\_9  
## loading 0.3261 0.5952 0.9172 0.2784   
## attribute\_2 0.8921 0.0776 0.6543 0.9381   
## attribute\_3 0.9896 0.1612 0.6937 0.7808   
## measurement\_0 0.5703 0.7408 0.2169 0.2048   
## measurement\_1 0.9024 0.0049 0.8668 0.4104   
## measurement\_2 0.1885 0.6511 0.5388 0.8064   
## measurement\_3 0.3255 0.3228 0.5652 0.2155   
## measurement\_4 0.7115 0.5452 0.0433 0.0935   
## measurement\_5 0.3329 0.2179 0.7561 0.7520   
## measurement\_6 0.2908 0.4506 0.1512   
## measurement\_7 0.2908 0.6621 0.6173   
## measurement\_8 0.4506 0.6621 0.4003   
## measurement\_9 0.1512 0.6173 0.4003   
## measurement\_10 0.9680 0.8647 0.4966 0.4307   
## measurement\_11 0.5956 0.2240 0.0597 0.4009   
## measurement\_12 0.2674 0.5056 0.8785 0.2466   
## measurement\_13 0.4617 0.5898 0.6189 0.3664   
## measurement\_14 0.8292 0.8961 0.4173 0.0571   
## measurement\_15 0.9801 0.9227 0.2634 0.6527   
## measurement\_16 0.7760 0.2732 0.3741 0.6322   
## measurement\_17 0.0000 0.0000 0.0000 0.0000   
## measurement\_10 measurement\_11 measurement\_12 measurement\_13  
## loading 0.2416 0.7794 0.8625 0.2841   
## attribute\_2 0.4044 0.0000 0.0000 0.0000   
## attribute\_3 0.0000 0.0000 0.0000 0.0000   
## measurement\_0 0.0000 0.6921 0.0000 0.0000   
## measurement\_1 0.0709 0.0000 0.0000 0.2525   
## measurement\_2 0.0037 0.7566 0.9376 0.0000   
## measurement\_3 0.9619 0.2761 0.5404 0.3393   
## measurement\_4 0.0364 0.9728 0.6448 0.8922   
## measurement\_5 0.7463 0.0562 0.6545 0.0420   
## measurement\_6 0.9680 0.5956 0.2674 0.4617   
## measurement\_7 0.8647 0.2240 0.5056 0.5898   
## measurement\_8 0.4966 0.0597 0.8785 0.6189   
## measurement\_9 0.4307 0.4009 0.2466 0.3664   
## measurement\_10 0.0000 0.0000 0.0000   
## measurement\_11 0.0000 0.0000 0.0000   
## measurement\_12 0.0000 0.0000 0.7641   
## measurement\_13 0.0000 0.0000 0.7641   
## measurement\_14 0.0000 0.0000 0.0005 0.5175   
## measurement\_15 0.0000 0.0000 0.0000 0.0042   
## measurement\_16 0.9127 0.0000 0.0000 0.0000   
## measurement\_17 0.5372 0.1366 0.2002 0.0027   
## measurement\_14 measurement\_15 measurement\_16 measurement\_17  
## loading 0.6975 0.3597 0.0538 0.0314   
## attribute\_2 0.0123 0.0000 0.0000 0.0000   
## attribute\_3 0.0000 0.0000 0.0000 0.0000   
## measurement\_0 0.0021 0.0000 0.5397 0.0006   
## measurement\_1 0.1050 0.0000 0.4497 0.1922   
## measurement\_2 0.0506 0.9660 0.0002 0.0052   
## measurement\_3 0.6333 0.1436 0.9992 0.0000   
## measurement\_4 0.9011 0.4171 0.4527 0.0000   
## measurement\_5 0.5417 0.7452 0.1927 0.0000   
## measurement\_6 0.8292 0.9801 0.7760 0.0000   
## measurement\_7 0.8961 0.9227 0.2732 0.0000   
## measurement\_8 0.4173 0.2634 0.3741 0.0000   
## measurement\_9 0.0571 0.6527 0.6322 0.0000   
## measurement\_10 0.0000 0.0000 0.9127 0.5372   
## measurement\_11 0.0000 0.0000 0.0000 0.1366   
## measurement\_12 0.0005 0.0000 0.0000 0.2002   
## measurement\_13 0.5175 0.0042 0.0000 0.0027   
## measurement\_14 0.0068 0.0000 0.5038   
## measurement\_15 0.0068 0.0193 0.6991   
## measurement\_16 0.0000 0.0193 0.0288   
## measurement\_17 0.5038 0.6991 0.0288

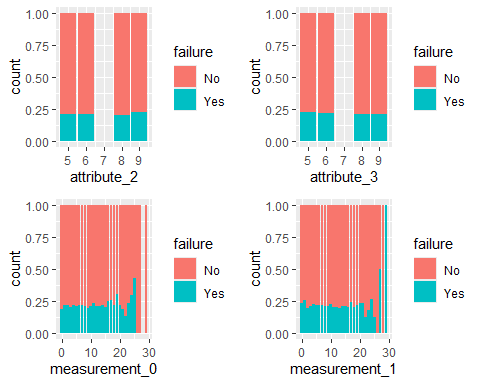
p1 = ggplot(train\_complete, aes(x = product\_code, fill = failure)) + geom\_bar(position = "fill")  
p2 = ggplot(train\_complete, aes(x = loading, fill = failure)) + geom\_histogram(position = "fill")  
p3 = ggplot(train\_complete, aes(x = attribute\_0, fill = failure)) + geom\_bar(position = "fill")  
p4 = ggplot(train\_complete, aes(x = attribute\_1, fill = failure)) + geom\_bar(position = "fill") + theme(axis.text = element\_text(size = 7))  
grid.arrange(p1,p2,p3,p4)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).



p5 = ggplot(train\_complete, aes(x = attribute\_2, fill = failure)) + geom\_bar(position = "fill")  
p6 = ggplot(train\_complete, aes(x = attribute\_3, fill = failure)) + geom\_bar(position = "fill")  
p7 = ggplot(train\_complete, aes(x = measurement\_0, fill = failure)) + geom\_bar(position = "fill")  
p8 = ggplot(train\_complete, aes(x = measurement\_1, fill = failure)) + geom\_bar(position = "fill")  
grid.arrange(p5,p6,p7,p8)



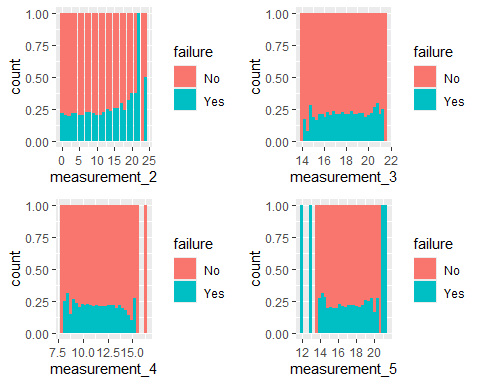
p9 = ggplot(train\_complete, aes(x = measurement\_2, fill = failure)) + geom\_bar(position = "fill")  
p10 = ggplot(train\_complete, aes(x = measurement\_3, fill = failure)) + geom\_histogram(position = "fill")  
p11 = ggplot(train\_complete, aes(x = measurement\_4, fill = failure)) + geom\_histogram(position = "fill")  
p12 = ggplot(train\_complete, aes(x = measurement\_5, fill = failure)) + geom\_histogram(position = "fill")  
grid.arrange(p9,p10,p11,p12)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 6 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).



p13 = ggplot(train\_complete, aes(x = measurement\_6, fill = failure)) + geom\_histogram(position = "fill")  
p14 = ggplot(train\_complete, aes(x = measurement\_7, fill = failure)) + geom\_histogram(position = "fill")  
p15 = ggplot(train\_complete, aes(x = measurement\_8, fill = failure)) + geom\_histogram(position = "fill")  
p16 = ggplot(train\_complete, aes(x = measurement\_9, fill = failure)) + geom\_histogram(position = "fill")  
grid.arrange(p13,p14,p15,p16)

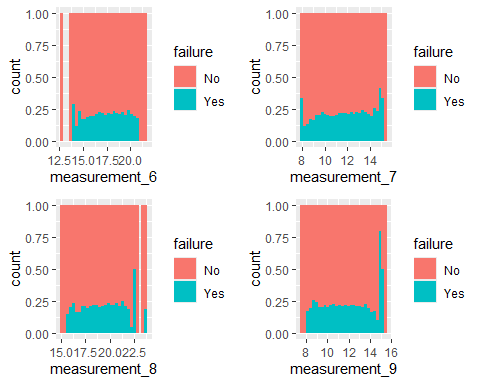
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



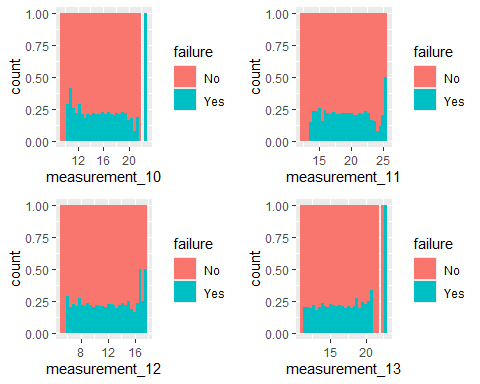
p17 = ggplot(train\_complete, aes(x = measurement\_10, fill = failure)) + geom\_histogram(position = "fill")  
p18 = ggplot(train\_complete, aes(x = measurement\_11, fill = failure)) + geom\_histogram(position = "fill")  
p19 = ggplot(train\_complete, aes(x = measurement\_12, fill = failure)) + geom\_histogram(position = "fill")  
p20 = ggplot(train\_complete, aes(x = measurement\_13, fill = failure)) + geom\_histogram(position = "fill")  
grid.arrange(p17,p18,p19,p20)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).



p21 = ggplot(train\_complete, aes(x = measurement\_14, fill = failure)) + geom\_histogram(position = "fill")  
p22 = ggplot(train\_complete, aes(x = measurement\_15, fill = failure)) + geom\_histogram(position = "fill")  
p23 = ggplot(train\_complete, aes(x = measurement\_16, fill = failure)) + geom\_histogram(position = "fill")  
p24 = ggplot(train\_complete, aes(x = measurement\_17, fill = failure)) + geom\_histogram(position = "fill")  
grid.arrange(p21,p22,p23,p24)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

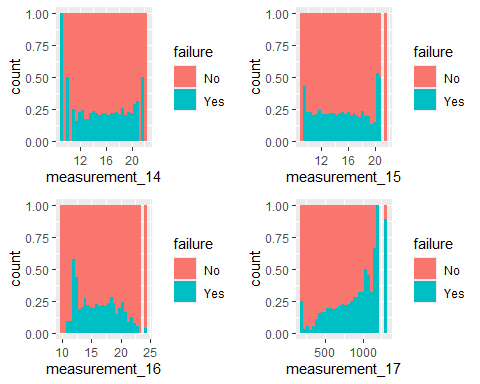
## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

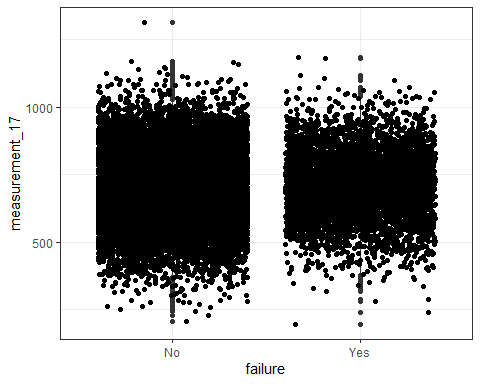
## Warning: Removed 4 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).



ggplot(train, aes(x=failure,y=measurement\_17)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 2284 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

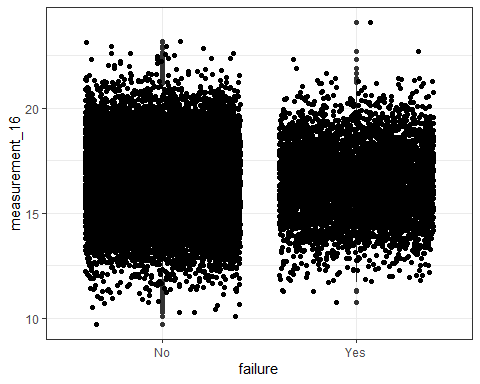
## Warning: Removed 2284 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_16)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 2110 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

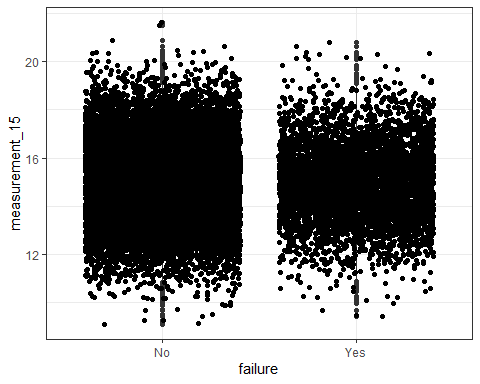
## Warning: Removed 2110 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_15)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 2009 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

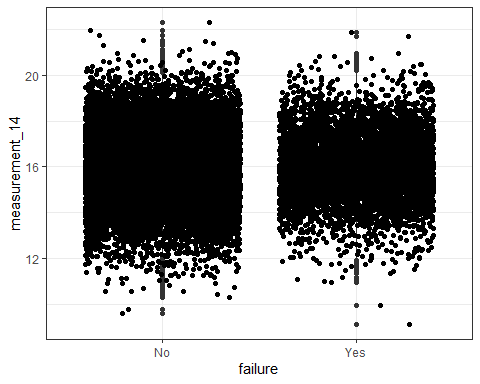
## Warning: Removed 2009 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_14)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 1874 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

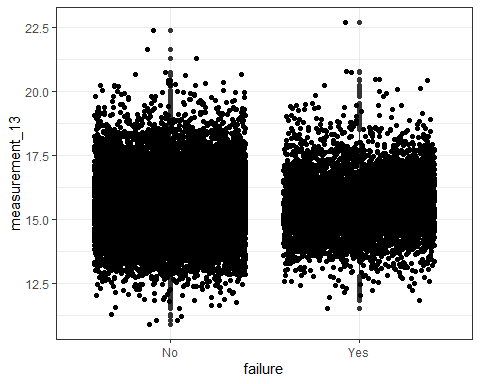
## Warning: Removed 1874 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_13)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 1774 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

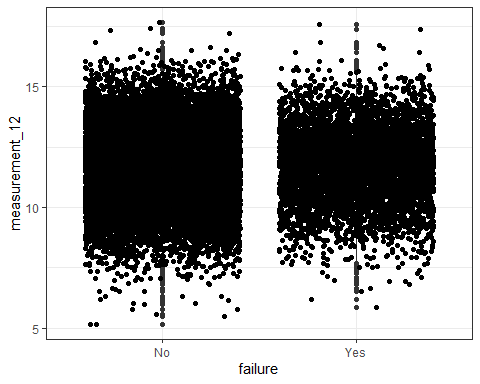
## Warning: Removed 1774 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_12)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 1601 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

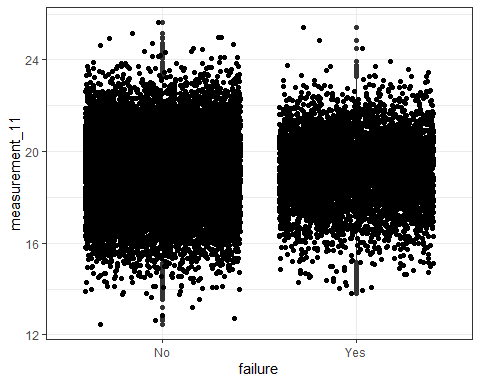
## Warning: Removed 1601 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_11)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 1468 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

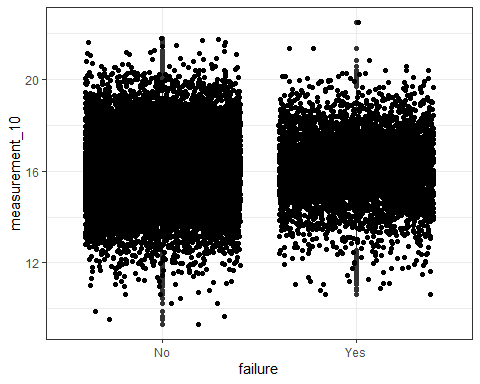
## Warning: Removed 1468 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_10)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 1300 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

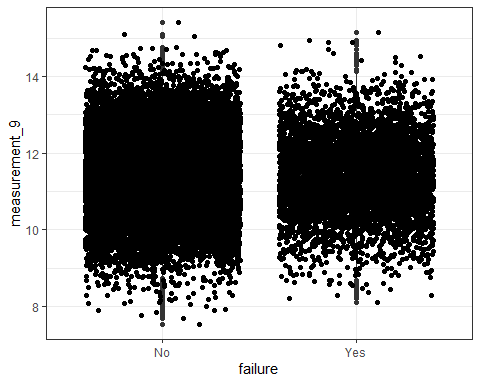
## Warning: Removed 1300 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_9)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 1227 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

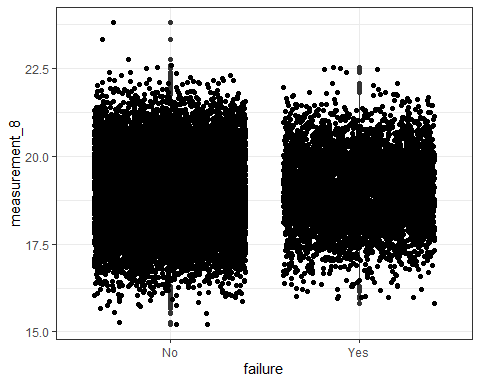
## Warning: Removed 1227 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_8)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 1048 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

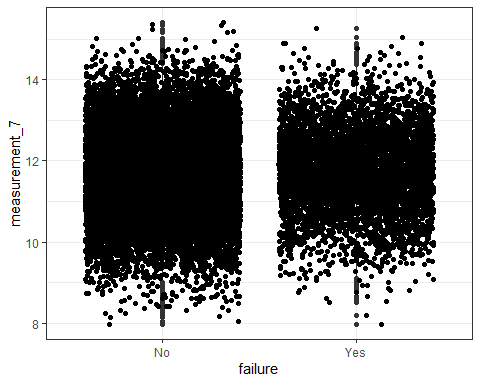
## Warning: Removed 1048 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_7)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 937 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

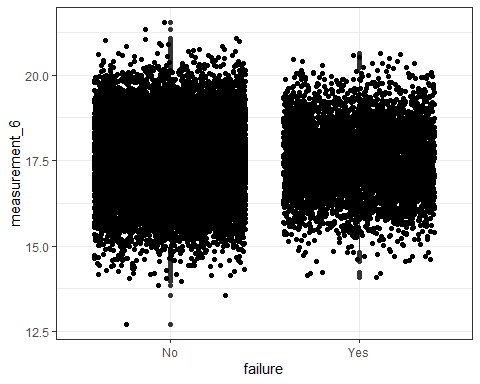
## Warning: Removed 937 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_6)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 796 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

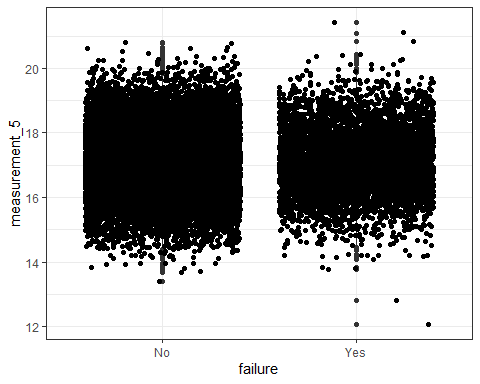
## Warning: Removed 796 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_5)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 676 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

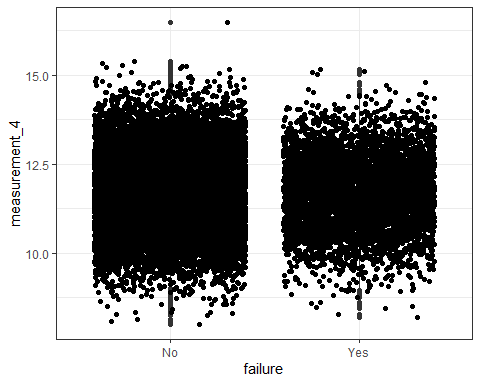
## Warning: Removed 676 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=measurement\_4)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 538 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

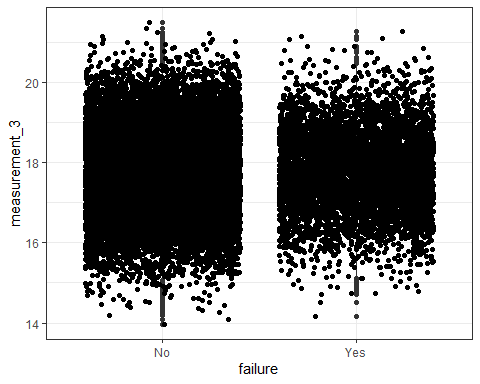
## Warning: Removed 538 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



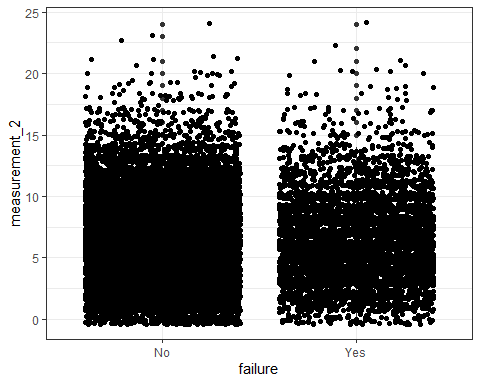
ggplot(train, aes(x=failure,y=measurement\_3)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 381 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

## Warning: Removed 381 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



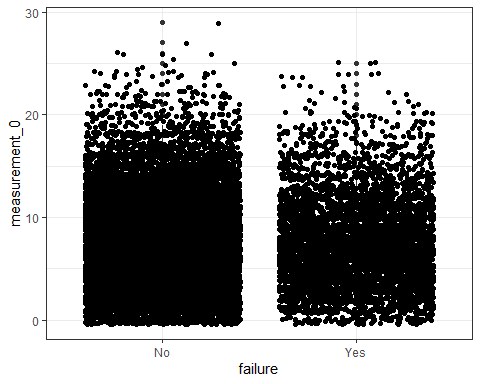
ggplot(train, aes(x=failure,y=measurement\_2)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()



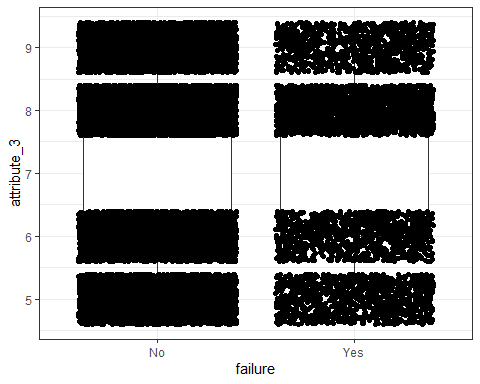
ggplot(train, aes(x=failure,y=measurement\_1)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()



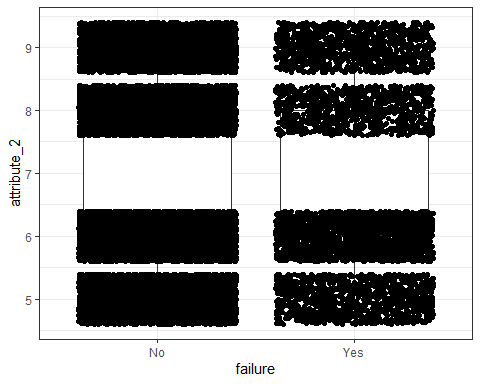
ggplot(train, aes(x=failure,y=measurement\_0)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()



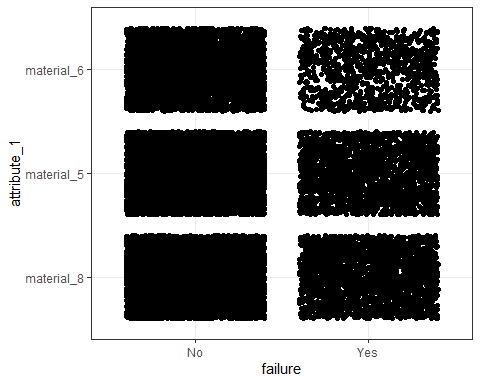
ggplot(train, aes(x=failure,y=attribute\_3)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()



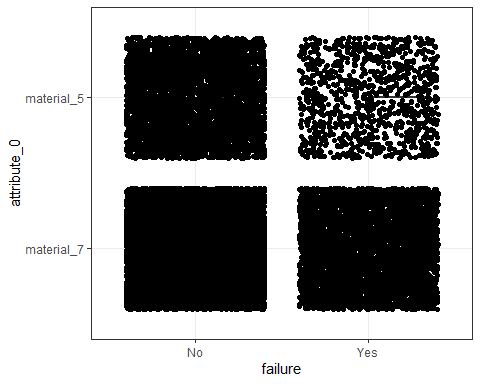
ggplot(train, aes(x=failure,y=attribute\_2)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()



ggplot(train, aes(x=failure,y=attribute\_1)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()



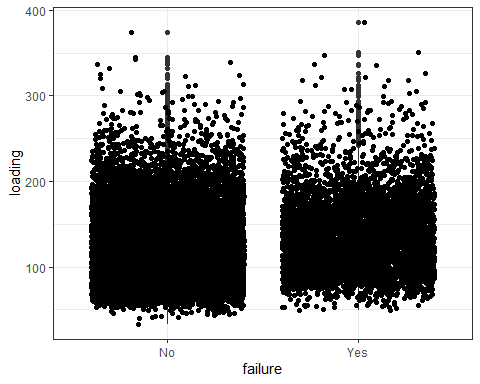
ggplot(train, aes(x=failure,y=attribute\_0)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()



ggplot(train, aes(x=failure,y=loading)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

## Warning: Removed 250 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).

## Warning: Removed 250 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(train, aes(x=failure,y=product\_code)) + geom\_boxplot() +   
 geom\_jitter() + theme\_bw()

