Investigating Missingness

Chris LeBoa

3/10/2021

## Investigating Missingness

## `summarise()` ungrouping output (override with `.groups` argument)

#Window Open Area

## # A tibble: 58 x 106  
## sampleid colldate hosp locationtype covidspace roomwidth roomlength  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl>  
## 1 02\_ICD1… 2020-10… icdd… Open ward COVID NA NA   
## 2 03\_SR1\_… 2020-10… Shei… Open ward COVID 6.93 13.9   
## 3 07\_SH1\_… 2020-10… Shah… Open ward COVID 13.5 16.1   
## 4 09\_KGH1… 2020-10… Kurm… Open ward COVID 7.78 21.4   
## 5 14\_KGH3… 2020-10… Kurm… Cabin COVID 9.30 3.51  
## 6 19\_KGH5… 2020-11… Kurm… Cabin COVID NA NA   
## 7 20\_SQ1\_… 2020-11… Squa… Cabin COVID 4.42 6.58  
## 8 21\_SQ1\_… 2020-11… Squa… Cabin COVID 7.09 7.80  
## 9 22\_SQ2\_… 2020-11… Squa… Open ward COVID NA NA   
## 10 23\_SQ2\_… 2020-11… Squa… Open ward COVID NA NA   
## # … with 48 more rows, and 99 more variables: roomheight <dbl>, roomvol <dbl>,  
## # floorarea <dbl>, wallarea <dbl>, surfacearea <dbl>, outdoorco2 <dbl>,  
## # co2start <dbl>, co2\_5min <dbl>, co2\_10min <dbl>, co2\_15min <dbl>,  
## # co2\_20min <dbl>, co2\_25min <dbl>, co2end <dbl>, fixco2 <dbl>,  
## # outdoorco2new <dbl>, co2startnew <dbl>, co2\_5minnew <dbl>,  
## # co2\_10minnew <dbl>, co2\_15minnew <dbl>, co2\_20minnew <dbl>,  
## # co2\_25minnew <dbl>, co2endnew <dbl>, co2average <dbl>, numcovidstart <dbl>,  
## # numnoncovidstart <dbl>, numstaffstart <dbl>, numotherstart <dbl>,  
## # nummasksstart <dbl>, numpeoplestart <dbl>, numcovidmid <dbl>,  
## # numnoncovidmid <dbl>, numstaffmid <dbl>, numothermid <dbl>,  
## # numpeoplemid <dbl>, numcovidend <dbl>, numnoncovidend <dbl>,  
## # numstaffend <dbl>, numotherend <dbl>, numpeopleend <dbl>,  
## # numpeopleavg <dbl>, popdensitystart <dbl>, popdensitymid <dbl>,  
## # popdensityend <dbl>, popdensityavg <dbl>, numnasal <dbl>, numoxymask <dbl>,  
## # numcpap <dbl>, numintub <dbl>, distpatient <dbl>, distwindow <dbl>,  
## # nearwindowopen <chr>, distdoor <dbl>, neardooropen <chr>,  
## # numwinfullopen <dbl>, numwinpartopen <dbl>, numwinclosed <dbl>,  
## # numwinopen <dbl>, openwinarea <dbl>, smallestopenwinarea <dbl>,  
## # numwintotal <dbl>, numdoorfullopen <dbl>, numdoorpartopen <dbl>,  
## # numdoorclosed <dbl>, numdoortotal <dbl>, numdooropen <dbl>,  
## # opendoorarea <dbl>, smallestopendoorarea <dbl>, totfan <dbl>,  
## # numfanon <dbl>, numacon <dbl>, tempstart <dbl>, humiditystart <dbl>,  
## # humidityend <dbl>, tempend <dbl>, wintofloorarea <dbl>,  
## # doortofloorarea <dbl>, windoortofloorarea <dbl>, ventratestart <dbl>,  
## # ventratemid <dbl>, ventrateend <dbl>, ventrateavg <dbl>, Q <dbl>,  
## # ach <dbl>, starttime <chr>, endtime <chr>, submittime <chr>, notes <chr>,  
## # sampletype <chr>, result <chr>, n1result <chr>, n1ctvalue <chr>,  
## # n1quantityreaction <dbl>, n1quantitysamplex10 <chr>, n2result <chr>,  
## # n2ctvalue <chr>, n2quantityreaction <dbl>, n2quantitysamplex10 <dbl>,  
## # batchnumber <chr>, rnadate <dbl>

For the most part this is NA for when the number of windows open is 0. In four instances there Is also an NA for NumWinOpen

We should discuss those four

## # A tibble: 4 x 3  
## sampleid colldate locationtype  
## <chr> <chr> <chr>   
## 1 02\_ICD1\_R2 2020-10-07 Open ward   
## 2 03\_SR1\_R1 2020-10-12 Open ward   
## 3 64\_KM2\_R6 2021-01-18 Doffing room  
## 4 78\_EC5\_R14 2021-02-08 OPD

#Door Open Area For the total door open area these should all be 0 since they all have 0 doors open

for icd1\_r1 there are a list of 3 open doors but the open door area parameter reads 0. Why?

## # A tibble: 19 x 1  
## numdooropen  
## <dbl>  
## 1 0  
## 2 0  
## 3 0  
## 4 0  
## 5 0  
## 6 0  
## 7 0  
## 8 0  
## 9 0  
## 10 0  
## 11 0  
## 12 0  
## 13 0  
## 14 0  
## 15 0  
## 16 0  
## 17 0  
## 18 0  
## 19 0

## # A tibble: 20 x 3  
## sampleid opendoorarea numdooropen  
## <chr> <dbl> <dbl>  
## 1 01\_ICD1\_R1 0 3  
## 2 02\_ICD1\_R2 NA 0  
## 3 05\_SR1\_R3 NA 0  
## 4 22\_SQ2\_R3 NA 0  
## 5 23\_SQ2\_R4 NA 0  
## 6 24\_SQ3\_R5 NA 0  
## 7 25\_SQ3\_R6 NA 0  
## 8 39\_DMC3\_R10 NA 0  
## 9 62\_KM2\_R4 NA 0  
## 10 64\_KM2\_R6 NA 0  
## 11 66\_EC1\_R2 NA 0  
## 12 68\_EC2\_R4 NA 0  
## 13 69\_EC2\_R5 NA 0  
## 14 70\_EC2\_R6 NA 0  
## 15 71\_EC3\_R7 NA 0  
## 16 72\_EC3\_R8 NA 0  
## 17 73\_EC3\_R9 NA 0  
## 18 76\_EC4\_R12 NA 0  
## 19 79\_EC5\_R15 NA 0  
## 20 87\_ICD2\_R6 NA 0

For the two that are missing temp end measurements we should add the temp that they started at

## # A tibble: 2 x 2  
## sampleid tempstart  
## <chr> <dbl>  
## 1 36\_DMC2\_R7 27.2  
## 2 58\_MM3\_R9 26.4

and for the sample missing a starting temperature we should add the ending temperature

## # A tibble: 1 x 2  
## sampleid tempend  
## <chr> <dbl>  
## 1 80\_EC5\_R16 22.6

For the window to floor area measure we should hand calculate sample 5, add in 0 to the open window area for rooms with no open windows

## # A tibble: 59 x 5  
## sampleid openwinarea numwintotal numwinopen floorarea  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 02\_ICD1\_R2 NA NA NA 567.   
## 2 03\_SR1\_R1 NA NA NA 96.3  
## 3 05\_SR1\_R3 1.01 1 1 21.5  
## 4 07\_SH1\_R2 NA 0 0 218.   
## 5 09\_KGH1\_R1 NA 0 0 166.   
## 6 14\_KGH3\_R6 NA 1 0 32.6  
## 7 19\_KGH5\_R10 NA 2 0 247.   
## 8 20\_SQ1\_R1 NA 2 0 29.1  
## 9 21\_SQ1\_R2 NA 0 0 55.2  
## 10 22\_SQ2\_R3 NA 1 0 914.   
## # … with 49 more rows

For sample missing a humidity at the start we add the humidity from the end and vice versa

## # A tibble: 1 x 2  
## sampleid humidityend  
## <chr> <dbl>  
## 1 80\_EC5\_R16 53.2

For the smallest open door area only the first sample had problems - says 3 doors but no area measure

Rest should be 0

## # A tibble: 20 x 4  
## sampleid numdooropen numdoorfullopen numdoortotal  
## <chr> <dbl> <dbl> <dbl>  
## 1 01\_ICD1\_R1 3 3 3  
## 2 02\_ICD1\_R2 0 0 2  
## 3 05\_SR1\_R3 0 0 1  
## 4 22\_SQ2\_R3 0 0 1  
## 5 23\_SQ2\_R4 0 0 1  
## 6 24\_SQ3\_R5 0 0 1  
## 7 25\_SQ3\_R6 0 0 0  
## 8 39\_DMC3\_R10 0 0 0  
## 9 62\_KM2\_R4 0 0 2  
## 10 64\_KM2\_R6 0 0 1  
## 11 66\_EC1\_R2 0 0 1  
## 12 68\_EC2\_R4 0 0 1  
## 13 69\_EC2\_R5 0 0 1  
## 14 70\_EC2\_R6 0 0 1  
## 15 71\_EC3\_R7 0 0 1  
## 16 72\_EC3\_R8 0 0 3  
## 17 73\_EC3\_R9 0 0 1  
## 18 76\_EC4\_R12 0 0 1  
## 19 79\_EC5\_R15 0 0 1  
## 20 87\_ICD2\_R6 0 0 2

There are no missing middle ventilation rates so we do not have to worry about replacement

## # A tibble: 8 x 4  
## sampleid ventratestart ventrateend ventrateavg  
## <chr> <dbl> <dbl> <dbl>  
## 1 01\_ICD1\_R1 13.6 10.8 12.2   
## 2 02\_ICD1\_R2 22.8 19.5 21.3   
## 3 28\_SQ4\_R9 18.1 19.5 18.8   
## 4 29\_SQ4\_R10 16.0 18.4 17.3   
## 5 36\_DMC2\_R7 5.05 NA 5.05  
## 6 64\_KM2\_R6 33.3 24.5 25.1   
## 7 67\_EC1\_R3 7.02 6.78 6.72  
## 8 74\_EC4\_R10 5.44 5.06 5.36

## # A tibble: 0 x 4  
## # … with 4 variables: sampleid <chr>, ventratestart <dbl>, ventrateend <dbl>,  
## # ventratemid <dbl>