Group 4 Report

Richard Davies, Katarina Lundervol, Pakwanja Desiree Twea

2024-09-16

TidyR data setup

The data is composed of two datasets with other lapping patients (see ID), exam.dataset.txt consists of PCR rdts for COVID19 during the 2020 pandemic, data includes information on patients including clinic, gender, test results week of pandemic, and exam_joindata.txt containing endpoint titer data from some of the patients (information can be found in the files codebook_exam_data.html, exam.descr.md). Data files were joined and processed as per conventions of tidyR for later visulization and analysis.

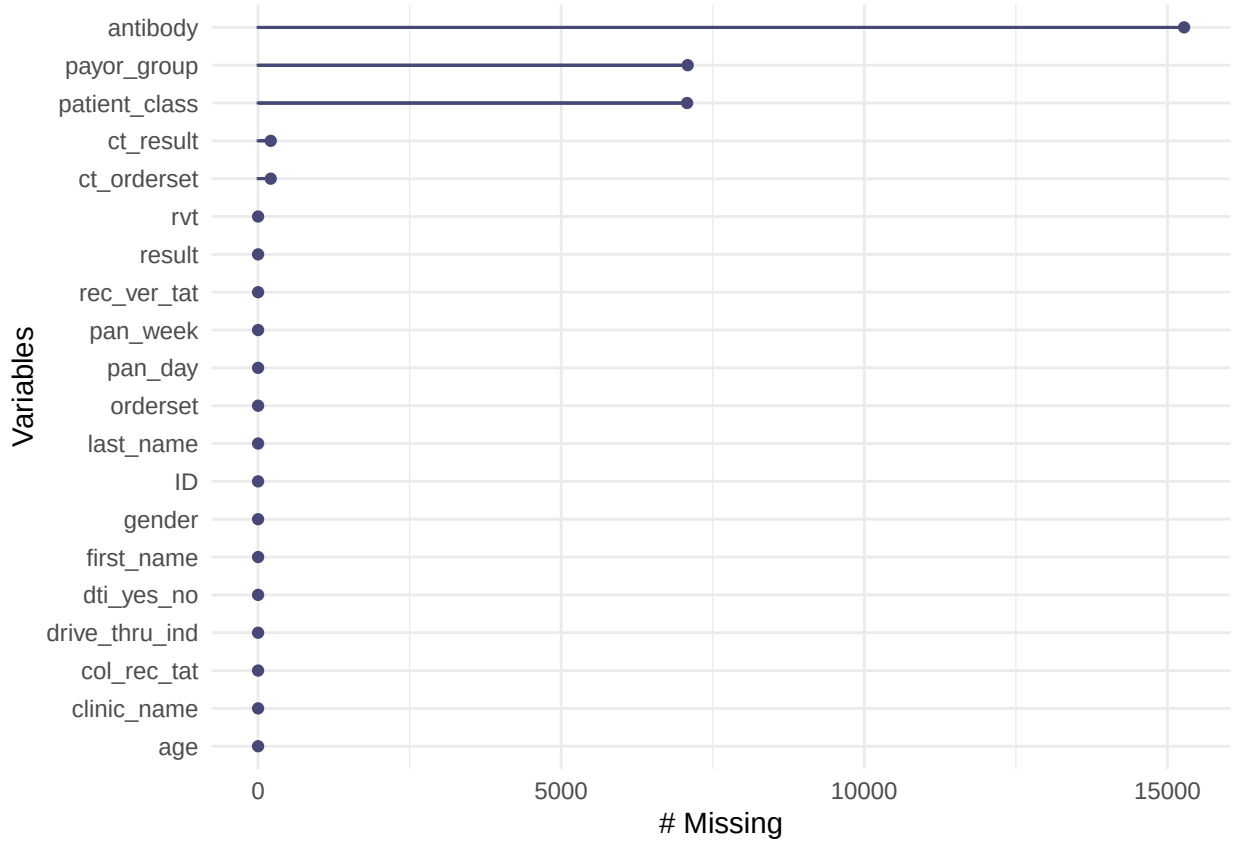
An overview of the data follows

```
summary(complete_data)
```

```
##        ID             age            gender      first_name
##  Min.   :    1   Min.   :  0.00   female:7832   Length:15524
##  1st Qu.: 2330   1st Qu.:  2.00   male  :7692   Class :character
##  Median : 5268   Median :  9.00                 Mode  :character
##  Mean   : 5571   Mean   : 14.19
##  3rd Qu.: 8636   3rd Qu.: 18.00
##  Max.   :12346   Max.   :138.00
##
##   last_name                    clinic_name        result      drive_thru_ind
##  Length:15524       clinical lab     :7500   invalid :  301   0:7537
##  Class :character   emergency dept   :3413   negative:14358   1:7987
##  Mode  :character   oncology day hosp: 533   positive:  865
##                     nicu             : 294
##                     laboratory       : 270
##                     picu             : 261
##                     (Other)          :3253
##    ct_result        orderset                  payor_group
##  Min.   :14.05   Min.   :0.0000   commercial        :3726
##  1st Qu.:45.00   1st Qu.:0.0000   government        :3644
##  Median :45.00   Median :1.0000   unassigned        : 733
##  Mean   :44.12   Mean   :0.6952   self pay          : 216
##  3rd Qu.:45.00   3rd Qu.:1.0000   medical assistance:  84
##  Max.   :45.00   Max.   :1.0000   (Other)           :  34
##  NA's   :209                      NA's              :7087
##              patient_class     pan_day        rec_ver_tat
##  inpatient           :3438   Min.   :  4.00   Min.   :-18.600
##  emergency           :1378   1st Qu.: 38.00   1st Qu.:  4.000
##  not applicable      :1096   Median : 65.00   Median :  5.000
##  outpatient          : 973   Mean   : 63.21   Mean   :  5.639
##  recurring outpatient: 795   3rd Qu.: 87.00   3rd Qu.:  6.200
```

```
##  (Other)             : 767   Max.   :107.00   Max.   :218.200
##  NA's                :7077
##   col_rec_tat           rvt              pan_week        dti_yes_no
##  Min.   :    0.00   Length:15524       Min.   : 0.5714   Length:15524
##  1st Qu.:    0.70   Class :character   1st Qu.: 5.4286   Class :character
##  Median :    1.90   Mode  :character   Median : 9.2857   Mode  :character
##  Mean   :    7.22                      Mean   : 9.0298
##  3rd Qu.:    3.60                      3rd Qu.:12.4286
##  Max.   :61370.20                      Max.   :15.2857
##
##   ct_orderset      antibody
##  Min.   : 0.0   Min.   : 21.48
##  1st Qu.: 0.0   1st Qu.: 63.30
##  Median :45.0   Median :103.52
##  Mean   :30.6   Mean   :109.99
##  3rd Qu.:45.0   3rd Qu.:156.95
##  Max.   :45.0   Max.   :199.89
##  NA's   :209    NA's   :15275
```

```
naniar::gg_miss_var(complete_data)
```

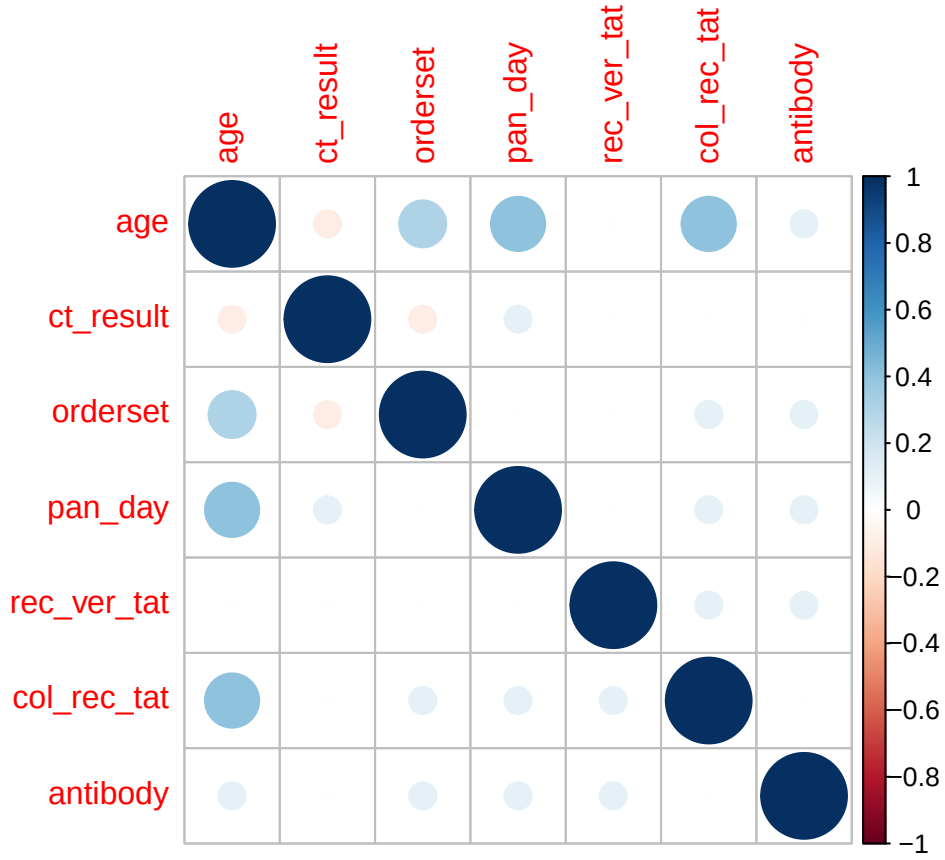


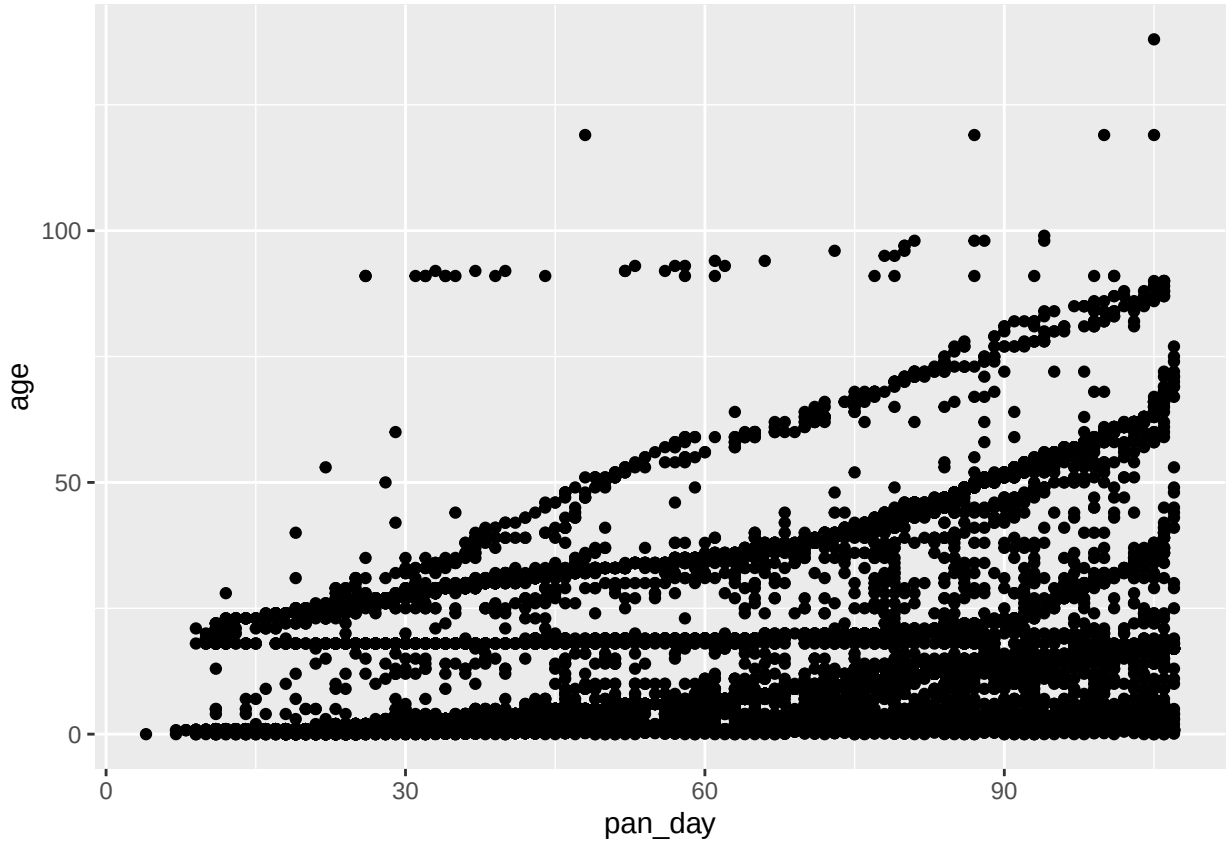
Descriptive plots

Visulization was done using a number of plots. Significant correlations are observiable with age, and orderset and days since the pandemic. Indicating early in the pandemic individuals seen in the clinic were usually

young adults and children

```
## # A tibble: 7 x 8
##   rowname           age ct_result  orderset   pan_day rec_ver_tat col_rec_tat
##   <chr>           <dbl>     <dbl>     <dbl>     <dbl>       <dbl>       <dbl>
## 1 age         0          4.54e-17 9.34e-151 3.95e-323    0.0666         0.844
## 2 ct_result   4.54e- 17  0        3   e- 11 7.86e-  1    0.0829         0.849
## 3 orderset    9.34e-151  3   e-11 0         7.38e-  1    0.000108       0.158
## 4 pan_day     3.95e-323  7.86e- 1 7.38e-  1 0            0.125          0.809
## 5 rec_ver_tat 6.66e-  2  8.29e- 2 1.08e-  4 1.25e-  1    0              0.854
## 6 col_rec_tat 8.44e-  1  8.49e- 1 1.58e-  1 8.09e-  1    0.854          0
## 7 antibody    4.16e-  2  9.53e- 1 2.89e-  1 4.05e-  2    0.0704         0.623
## # i 1 more variable: antibody <dbl>
```

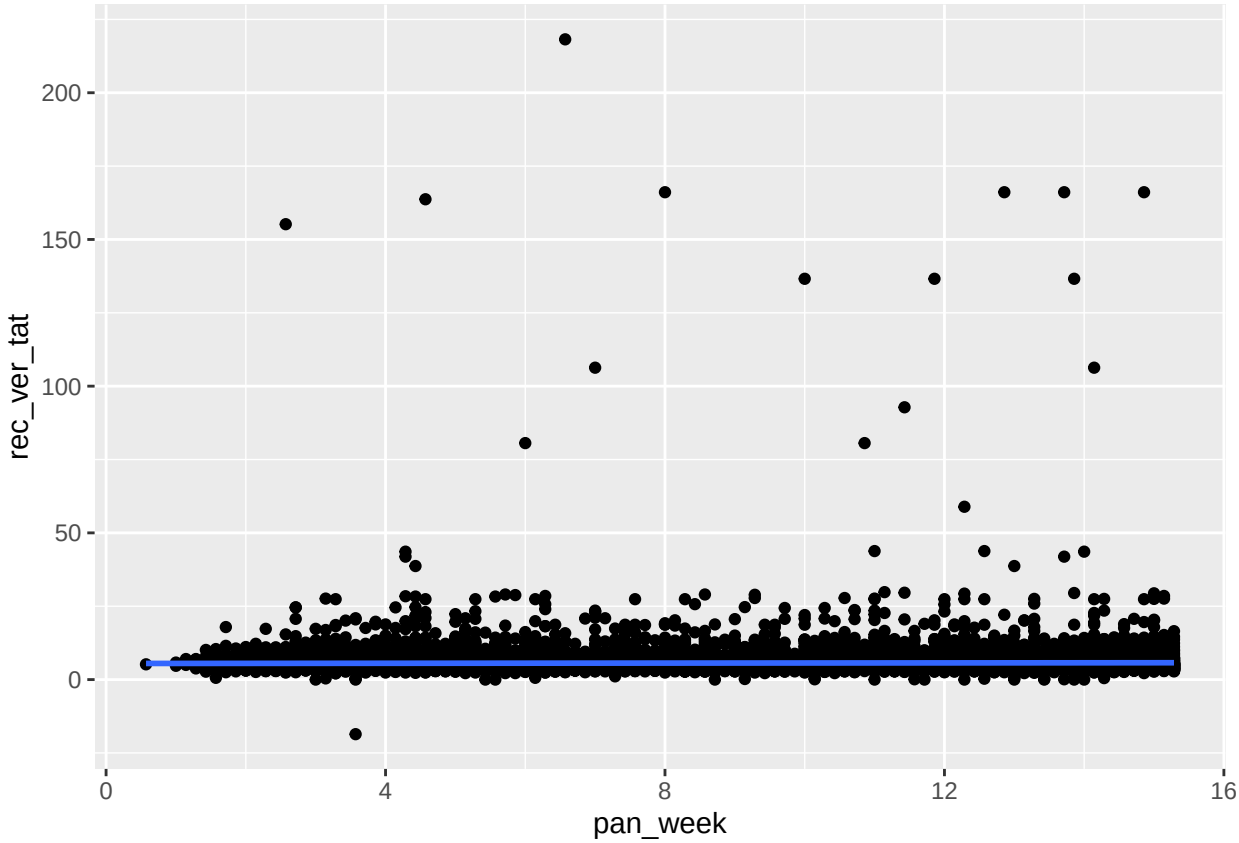




Suprisingly over the course of the pandemic there was no indication that the wait time from collection of patient sample to PCR test result was reduced.

```
## # A tibble: 15 x 4
##    pan_week mean_value   std `median(rec_ver_tat)`
##       <dbl>      <dbl> <dbl>                 <dbl>
##  1        1       5.51  1.57                   5.3
##  2        2       5.27  2.00                   4.8
##  3        3       5.62  5.44                   5
##  4        4       5.50  3.19                   5
##  5        5       5.57  5.07                   5
##  6        6       5.69  3.61                   5.1
##  7        7       5.73  7.81                   4.9
##  8        8       5.66  5.97                   4.9
##  9        9       5.48  2.59                   4.9
## 10       10       5.65  4.68                   5
## 11       11       5.73  4.17                   5
## 12       12       5.57  4.73                   4.9
## 13       13       5.59  5.22                   4.9
## 14       14       5.82  6.88                   5
## 15       15       5.81  5.62                   5
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Statistics

wilcox test indicated that the more tests from a single patient does not associate with increase positive test, in fact the opposite trend was noted.

```
## # A tibble: 1 x 4
##   statistic  p.value method                                          alternative
##       <dbl>    <dbl> <chr>                                           <chr>
## 1   4979785 8.14e-11 Wilcoxon rank sum test with continuity correct~ two.sided
```

While drive-through use was associated with an increase in positive tests

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  positive_visits$n and positive_visits$positive
## X-squared = 44.749, df = NA, p-value = 0.01249
```