Zeyd Khalil HW7, October 8, 2020

Exercise 1

```
parse_date("01, January 2018", format = "%d, %B %Y")
```

```
## [1] "2018-01-01"
```

```
parse_date("01-January/2000", format = "%d-%B/%Y")
```

```
## [1] "2000-01-01"
```

```
parse_date("1 Jan 19", format = "%d %b %y")
```

```
## [1] "2019-01-01"
```

Exercise 2

```
parse_time("10:40 pm", format = "%I:%M %p")
```

```
## 22:40:00
```

```
parse_time("23:40-22", format = "%H:%M-%S")
```

```
## 23:40:22
```

Exercise 3

```
covidstats <- read_csv("owid-covid-data.csv", col_types = cols(date = col_date()))

head(covidstats)
```

```
## # A tibble: 6 x 41
##   iso_code continent location date       total_cases new_cases new_cases_smoot~
##   <chr>    <chr>     <chr>    <date>           <dbl>     <dbl>            <dbl>
## 1 AFG      Asia      Afghani~ 2019-12-31           0         0               NA
## 2 AFG      Asia      Afghani~ 2020-01-01           0         0               NA
```

```
## 3 AFG      Asia      Afghani~ 2020-01-02           0         0               NA
## 4 AFG      Asia      Afghani~ 2020-01-03           0         0               NA
## 5 AFG      Asia      Afghani~ 2020-01-04           0         0               NA
## 6 AFG      Asia      Afghani~ 2020-01-05           0         0               NA
## # ... with 34 more variables: total_deaths <dbl>, new_deaths <dbl>,
## #   new_deaths_smoothed <dbl>, total_cases_per_million <dbl>,
## #   new_cases_per_million <dbl>, new_cases_smoothed_per_million <dbl>,
## #   total_deaths_per_million <dbl>, new_deaths_per_million <dbl>,
## #   new_deaths_smoothed_per_million <dbl>, new_tests <lgl>, total_tests <lgl>,
## #   total_tests_per_thousand <lgl>, new_tests_per_thousand <lgl>,
## #   new_tests_smoothed <lgl>, new_tests_smoothed_per_thousand <lgl>,
## #   tests_per_case <lgl>, positive_rate <lgl>, tests_units <lgl>,
## #   stringency_index <dbl>, population <dbl>, population_density <dbl>,
## #   median_age <dbl>, aged_65_older <dbl>, aged_70_older <dbl>,
## #   gdp_per_capita <dbl>, extreme_poverty <dbl>, cardiovasc_death_rate <dbl>,
## #   diabetes_prevalence <dbl>, female_smokers <dbl>, male_smokers <dbl>,
## #   handwashing_facilities <dbl>, hospital_beds_per_thousand <dbl>,
## #   life_expectancy <dbl>, human_development_index <dbl>
```

```
summary(covidstats)
```

```
##    iso_code          continent           location              date
##  Length:44363       Length:44363       Length:44363       Min.   :2019-12-31
##  Class :character   Class :character   Class :character   1st Qu.:2020-04-12
##  Mode  :character   Mode  :character   Mode  :character   Median :2020-06-04
##                                                           Mean   :2020-06-01
##                                                           3rd Qu.:2020-07-27
##                                                           Max.   :2020-09-17
##
##   total_cases         new_cases      new_cases_smoothed   total_deaths
##  Min.   :       0   Min.   : -8261   Min.   :  -552.00   Min.   :     0
##  1st Qu.:      51   1st Qu.:     0   1st Qu.:     0.57   1st Qu.:     0
##  Median :     911   Median :     9   Median :    13.86   Median :    17
##  Mean   :   91847   Mean   :  1373   Mean   :  1358.67   Mean   :  3836
##  3rd Qu.:    9762   3rd Qu.:   154   3rd Qu.:   158.89   3rd Qu.:   208
##  Max.   :29902487   Max.   :310156   Max.   :286047.71   Max.   :941291
##  NA's   :603        NA's   :801      NA's   :1583        NA's   :603
##    new_deaths       new_deaths_smoothed total_cases_per_million
##  Min.   :-1918.00   Min.   :-232.143    Min.   :    0.0
##  1st Qu.:    0.00   1st Qu.:   0.000    1st Qu.:   24.1
##  Median :    0.00   Median :   0.143    Median :  264.6
##  Mean   :   43.22   Mean   :  43.264    Mean   : 1748.0
##  3rd Qu.:    3.00   3rd Qu.:   2.857    3rd Qu.: 1756.1
##  Max.   :10467.00   Max.   :7459.143    Max.   :42501.4
##  NA's   :801        NA's   :1583        NA's   :865
##  new_cases_per_million new_cases_smoothed_per_million total_deaths_per_million
##  Min.   :-2212.545     Min.   :-269.978               Min.   :   0.000
##  1st Qu.:    0.000     1st Qu.:   0.120               1st Qu.:   0.000
##  Median :    1.447     Median :   2.805               Median :   4.545
##  Mean   :   23.273     Mean   :  22.917               Mean   :  54.710
##  3rd Qu.:   16.220     3rd Qu.:  17.899               3rd Qu.:  33.072
##  Max.   : 4944.376     Max.   : 882.924               Max.   :1237.551
##  NA's   :865           NA's   :1648                   NA's   :865
##  new_deaths_per_million new_deaths_smoothed_per_million new_tests
```

```
##  Min.   :-67.9010       Min.   :-9.6780                 Mode:logical
##  1st Qu.:  0.0000       1st Qu.: 0.0000                 NA's:44363
##  Median :  0.0000       Median : 0.0180
##  Mean   :  0.5601       Mean   : 0.5603
##  3rd Qu.:  0.1960       3rd Qu.: 0.2740
##  Max.   :215.3820       Max.   :63.1400
##  NA's   :865            NA's   :1648
##  total_tests    total_tests_per_thousand new_tests_per_thousand
##  Mode:logical   Mode:logical             Mode:logical
##  NA's:44363     NA's:44363               NA's:44363
##
##
##
##
##
##  new_tests_smoothed new_tests_smoothed_per_thousand tests_per_case
##  Mode:logical       Mode:logical                    Mode:logical
##  NA's:44363         NA's:44363                      NA's:44363
##
##
##
##
##
##  positive_rate  tests_units    stringency_index   population
##  Mode:logical   Mode:logical   Min.   :  0.00   Min.   :8.090e+02
##  NA's:44363     NA's:44363     1st Qu.: 38.89   1st Qu.:1.403e+06
##                                Median : 64.35   Median :8.655e+06
##                                Mean   : 57.70   Mean   :8.927e+07
##                                3rd Qu.: 79.63   3rd Qu.:3.107e+07
##                                Max.   :100.00   Max.   :7.795e+09
##                                NA's   :7636     NA's   :262
##  population_density    median_age    aged_65_older    aged_70_older
##  Min.   :    0.137   Min.   :15.10   Min.   : 1.144   Min.   : 0.526
##  1st Qu.:   39.497   1st Qu.:23.50   1st Qu.: 3.556   1st Qu.: 2.142
##  Median :   88.125   Median :31.40   Median : 6.986   Median : 4.419
##  Mean   :  361.644   Mean   :31.37   Mean   : 9.285   Mean   : 5.875
##  3rd Qu.:  214.243   3rd Qu.:39.70   3rd Qu.:14.762   3rd Qu.: 9.473
##  Max.   :19347.500   Max.   :48.20   Max.   :27.049   Max.   :18.493
##  NA's   :2259        NA's   :4783    NA's   :5369     NA's   :4988
##  gdp_per_capita     extreme_poverty cardiovasc_death_rate diabetes_prevalence
##  Min.   :   661.2   Min.   : 0.10   Min.   : 79.37        Min.   : 0.990
##  1st Qu.:  5338.4   1st Qu.: 0.50   1st Qu.:153.51        1st Qu.: 5.310
##  Median : 14103.5   Median : 1.80   Median :238.34        Median : 7.110
##  Mean   : 20996.0   Mean   :12.01   Mean   :251.30        Mean   : 8.049
##  3rd Qu.: 32415.1   3rd Qu.:16.00   3rd Qu.:318.95        3rd Qu.:10.180
##  Max.   :116935.6   Max.   :77.60   Max.   :724.42        Max.   :23.360
##  NA's   :5287       NA's   :18285   NA's   :4768          NA's   :3383
##  female_smokers   male_smokers   handwashing_facilities
##  Min.   : 0.10   Min.   : 7.70   Min.   : 1.188
##  1st Qu.: 1.90   1st Qu.:21.40   1st Qu.:21.222
##  Median : 6.40   Median :31.40   Median :55.182
##  Mean   :10.84   Mean   :32.64   Mean   :52.598
##  3rd Qu.:19.60   3rd Qu.:40.90   3rd Qu.:83.741
##  Max.   :44.00   Max.   :78.10   Max.   :98.999
```

```
##  NA's   :13287   NA's   :13678   NA's   :25889
##  hospital_beds_per_thousand life_expectancy human_development_index
##  Min.   : 0.100             Min.   :53.28   Min.   :0.354
##  1st Qu.: 1.300             1st Qu.:69.91   1st Qu.:0.606
##  Median : 2.500             Median :75.49   Median :0.755
##  Mean   : 3.116             Mean   :74.06   Mean   :0.726
##  3rd Qu.: 4.200             3rd Qu.:79.93   3rd Qu.:0.853
##  Max.   :13.800             Max.   :86.75   Max.   :0.953
##  NA's   :8554               NA's   :814     NA's   :6163
```

```
write_csv(covidstats, "Data/owid-covid-data.csv", append = TRUE, na = "NA")
```

As seen in covidstats, there are several columns that are . These columns have no values in them, which therefore means I will be ignoring them for my project.

COVID-19 dataset is a collection of the COVID-19 data maintained by Our World in Data. It is updated daily and includes data on confirmed cases, deaths, and testing, as well as other variables of potential interest.