Assignment 2 - AS2018443

```
> install.packages("devtools")
Installing package into ‘/cloud/lib/x86_64-pc-linux-gnu-library/4.1’
(as ‘lib’ is unspecified)
trying URL 'http://package-proxy/focal/src/contrib/devtools_2.4.2.tar.gz'
Content type 'application/x-tar' length 396931 bytes (387 KB)
==================================================
downloaded 387 KB

* installing *binary* package ‘devtools’ ...
* DONE (devtools)

The downloaded source packages are in
        ‘/tmp/Rtmp9WhC2I/downloaded_packages’
> library(devtools)
Loading required package: usethis
> devtools::install_github("thiyangt/sta3262")
Skipping install of 'sta3262' from a github remote, the SHA1 (0e75c263) has not c
hanged since last install.
  Use `force = TRUE` to force installation
> library(sta3262)
> data<-get_assignment_data("AS2018443")
> data
# A tibble: 1,476 × 4
# Groups:   country, date [492]
   country date       type      cases
   <chr>   <date>     <chr>     <dbl>
 1 Canada  2020-01-22 confirmed     0
 2 Canada  2020-01-22 death         0
 3 Canada  2020-01-22 recovered     0
 4 Canada  2020-01-23 confirmed     0
 5 Canada  2020-01-23 death         0
 6 Canada  2020-01-23 recovered     0
 7 Canada  2020-01-24 confirmed     0
 8 Canada  2020-01-24 death         0
 9 Canada  2020-01-24 recovered     0
10 Canada  2020-01-25 confirmed     0
# … with 1,466 more rows
> get_assignment_questions(2018443)
$q1
[1] "Obtain summary statistics for each variable and interpret the results."

$q2
[1] "Draw a time series plot using the qplotfunction to visualize changes in Covi
d-19 recovered cases over time."

$q3
```

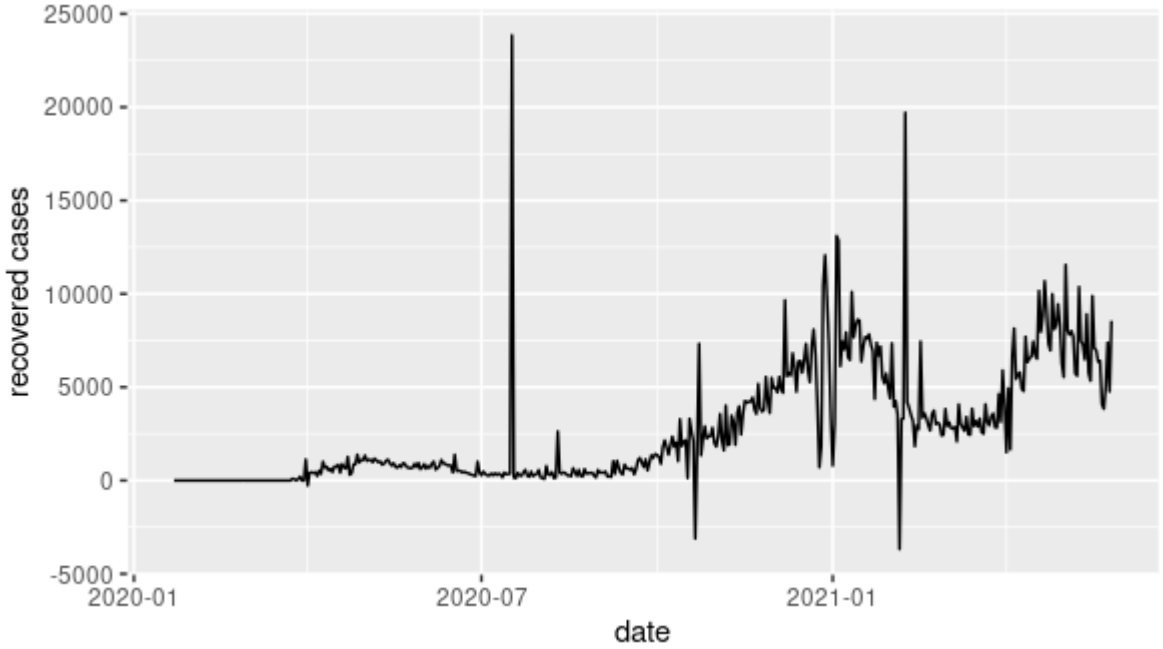
```
[1] "Draw a scatterplot using the qplot to visualize the relationship between Cov
id-19 confirmed cases and recovered cases. Compute the corresponding Pearson’s co
rrelation coefficient."

$q4
[1] "In the year 2021, which date has the highest number of recovered cases?"

$q5
[1] "In 2021, which dates are almost the same in the number of Covid-19 recovered
cases?"

$q6
[1] "Create a new dataframe called “newcovid”, that contains rows of only death c
ases."
```

```
> #Question1
> #"Obtain summary statistics for each variable and interpret the results."
> summary(data)
   country               date                type               cases
 Length:1476        Min.   :2020-01-22   Length:1476        Min.   :-3673
 Class :character   1st Qu.:2020-05-23   Class :character   1st Qu.:   34
 Mode  :character   Median :2020-09-23   Mode  :character   Median :  399
                    Mean   :2020-09-23                      Mean   : 1841
                    3rd Qu.:2021-01-24                      3rd Qu.: 2932
                    Max.   :2021-05-27                      Max.   :23848
> #Question2
> # "Draw a time series plot using the qplotfunction to visualize changes in Covi
d-19 recovered cases over time."
> recovered <- data[data$type=="recovered",]
> library(ggplot2)
```

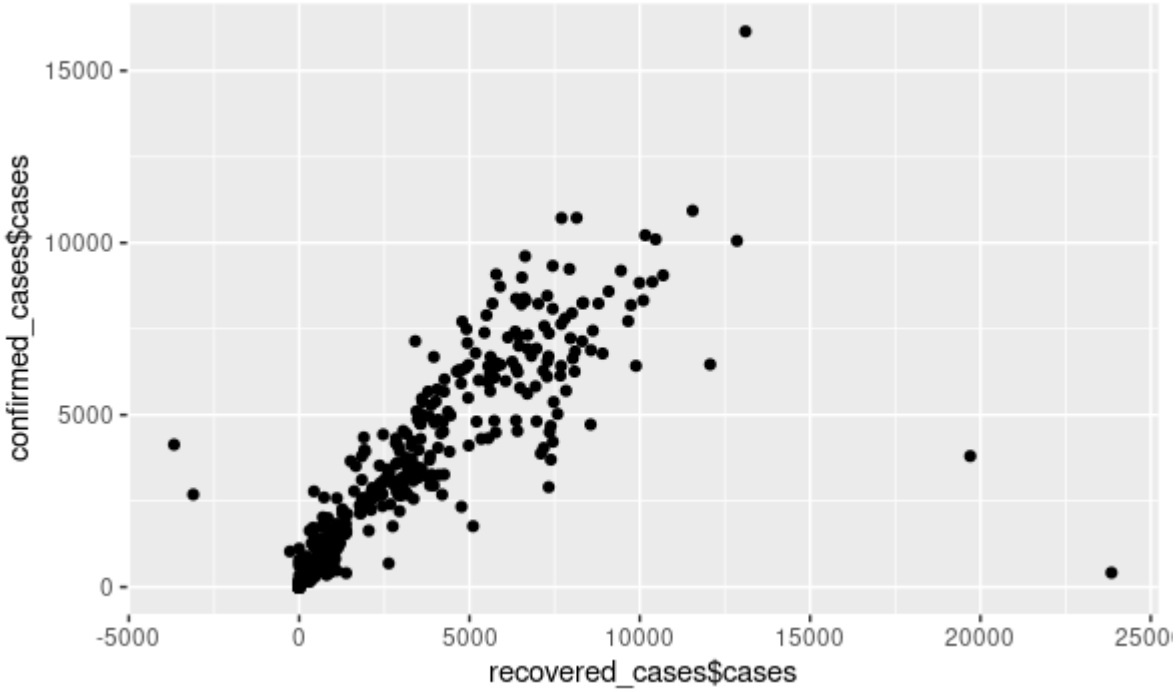



```
> qplot(data=recovered, y=cases, x=date, geom="line") + ylab("recovered cases")
> install.packages("dplyr")
```

```
Error in install.packages : Updating loaded packages
> install.packages("dplyr")
Installing package into ‘/cloud/lib/x86_64-pc-linux-gnu-library/4.1’
(as ‘lib’ is unspecified)
trying URL 'http://package-proxy/focal/src/contrib/dplyr_1.0.7.tar.gz'
Content type 'application/x-tar' length 1285014 bytes (1.2 MB)
==================================================
downloaded 1.2 MB

* installing *binary* package ‘dplyr’ ...
* DONE (dplyr)

The downloaded source packages are in
        ‘/tmp/Rtmp9WhC2I/downloaded_packages’
> #Question3
> confirmed_cases<- data[data$type=="confirmed",]
> recovered_cases<- data[data$type=="recovered",]
> Splot<-qplot(recovered_cases$cases, confirmed_cases$cases) #1
> Splot
```




```
> coef<-cor(recovered_cases$cases,confirmed_cases$cases) #2
> coef
[1] 0.8388935
> year2021<- data[data$date>="2021-01-01"& data$date<="2021-12-31",]
> year2021
# A tibble: 441 × 4
# Groups:   country, date [147]
   country date       type      cases
   <chr>   <date>     <chr>     <dbl>
 1 Canada  2021-01-01 confirmed   991
 2 Canada  2021-01-01 death        44
```

```
 3 Canada  2021-01-01 recovered   799
 4 Canada  2021-01-02 confirmed  4535
 5 Canada  2021-01-02 death       108
 6 Canada  2021-01-02 recovered  3055
 7 Canada  2021-01-03 confirmed 16141
 8 Canada  2021-01-03 death        92
 9 Canada  2021-01-03 recovered 13103
10 Canada  2021-01-04 confirmed 10058
# … with 431 more rows
> recoveredcases2021<-year2021[year2021$type=="recovered",]
> which(recoveredcases2021$cases==max(recoveredcases2021$cases))
[1] 39
> recoveredcases2021[62,]
# A tibble: 1 × 4
# Groups:   country, date [1]
  country date       type      cases
  <chr>   <date>     <chr>     <dbl>
1 Canada  2021-03-03 recovered  3115
> #Question5
> library(dplyr)
> recoveredcases2021%>%group_by(cases)%>%filter(n()>1)%>%summarize(n=n())
# A tibble: 0 × 2
# … with 2 variables: cases <dbl>, n <int>
> #Question6
> newcovid<- data[data$type=="death",]
> print(newcovid)
# A tibble: 492 × 4
# Groups:   country, date [492]
   country date       type  cases
   <chr>   <date>     <chr> <dbl>
 1 Canada  2020-01-22 death     0
 2 Canada  2020-01-23 death     0
 3 Canada  2020-01-24 death     0
 4 Canada  2020-01-25 death     0
 5 Canada  2020-01-26 death     0
 6 Canada  2020-01-27 death     0
 7 Canada  2020-01-28 death     0
 8 Canada  2020-01-29 death     0
 9 Canada  2020-01-30 death     0
10 Canada  2020-01-31 death     0
# … with 482 more rows
```