Question2_Set1_2017CSC1061.R

nitish

2020-06-05

```
# QUESTION 2
# RESEARCH PROBLEM ----
# ANALYSE age vs gender variations for different countries

# ALL EXTERNAL PACKAGES ----
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
# Setting Directory and Reading csv
# Filling NA in place of blanks in the cell ----
setwd("/home/nitish/Desktop/R_stuff/DataSc/PracAsgn")
covid <- read.csv("Covid_dataset2020.csv", header = TRUE, na.strings = c("",NA))
as_tibble(covid)
```

```
## # A tibble: 643 x 18
##    unique_id country case_in_country reporting_date sub_country gender   age
##        <int> <fct>              <int> <fct>          <fct>       <fct>  <int>
##  1         1 China                 NA 01/27/20       Hubei       female    28
##  2         2 China                 NA 01/27/20       Hubei       female    51
##  3         3 China                 NA 01/27/20       Shandong    male      37
##  4         4 Japan                  1 01/15/20       Kanagawa    male      35
##  5         5 Japan                  5 01/28/20       Aichi Pref~ male      45
##  6         6 Japan                  6 01/28/20       Nara Prefe~ male      65
##  7         7 Japan                  7 01/28/20       Hokkaido    female    45
##  8         8 Japan                  8 01/29/20       Osaka Pref~ female    45
```

```
##  9         9 Japan                  9 01/30/20       Tokyo       male      55
## 10        10 Japan                 10 01/30/20       Mie         male      55
## # ... with 633 more rows, and 11 more variables: symptom_onset <fct>,
## #   visit_date_hosp <fct>, intl_traveler <fct>, dom_traveler <int>,
## #   exposure_startdate <fct>, exposure_enddate <fct>, visiting_Wuhan <fct>,
## #   lives_in_Wuhan <fct>, death <fct>, recovered <fct>, symptom <fct>
```

```
#Formatting all date columns of the format mm/dd/yyyy(total 7 columns) ----
covid2 <- covid %>%
  mutate_at(vars(reporting_date, symptom_onset, visit_date_hosp,
         exposure_startdate, exposure_enddate, death, recovered), mdy)
```

```
## Warning: 26 failed to parse.
```

```
## Warning: 10 failed to parse.
```

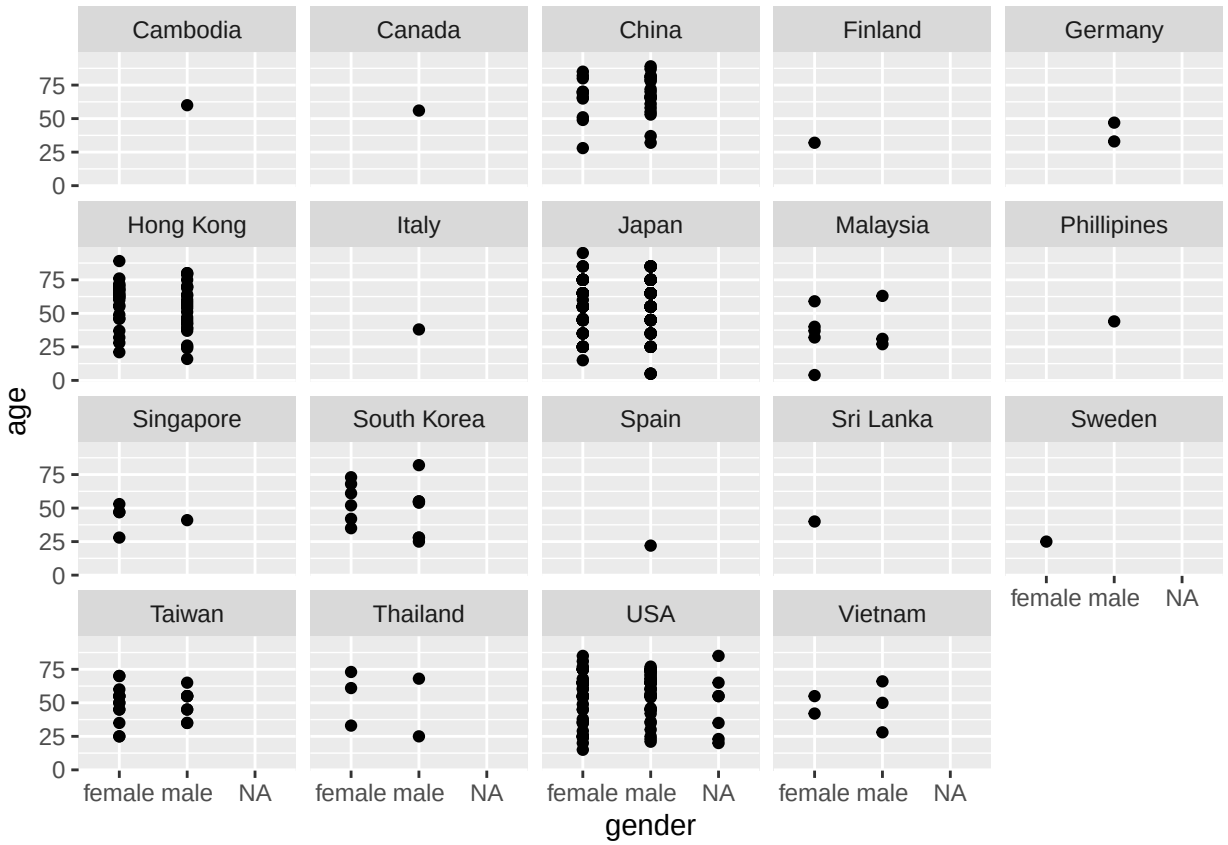
```
## Warning: 5 failed to parse.
```

```
## Warning: 442 failed to parse.
```

```
## Warning: 404 failed to parse.
```

```
write.csv(covid2,  "CleanData.csv")
#Grouping by country to create a facet map depicting various parameters ----
covid %>%
  group_by(country) %>%
  select(country, case_in_country, age, gender) %>%
  filter_all(any_vars(!is.na(.))) %>%
  ggplot() +
  geom_point(mapping = aes(x = gender, y = age)) +
  facet_wrap(vars(country))
```

```
## Warning: Removed 185 rows containing missing values (geom_point).
```




```
# Variance of case_in_country with age for each country ----
covid %>%
  group_by(country) %>%
  select(country, case_in_country, age, gender) %>%
  filter_all(any_vars(!is.na(.))) %>%
  summarise(correlation_age_cases = var(age, case_in_country, na.rm = TRUE))
```

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

```
## # A tibble: 19 x 2
##    country     correlation_age_cases
##    <fct>                       <dbl>
##  1 Cambodia                     NA
##  2 Canada                       NA
##  3 China                        NA
##  4 Finland                      NA
##  5 Germany                     140
##  6 Hong Kong                    34.0
##  7 Italy                        NA
##  8 Japan                         9.86
##  9 Malaysia                     13.5
## 10 Phillipines                  NA
## 11 Singapore                   -52.8
## 12 South Korea                 102.
## 13 Spain                        NA
## 14 Sri Lanka                    NA
## 15 Sweden                       NA
## 16 Taiwan                      -62.6
```

```
## 17 Thailand                 -270
## 18 USA                       119.
## 19 Vietnam                     9.85
```