Data Science Lab - Assignment 1

Lorenzo Mauri

6 maggio 2020

```
library(magrittr) #package for concatenating commands %>%
library(tibble)  #set data as tibble object
library(dplyr)  # data management (filter,select,summarize,ecc...)
```

```
##
## 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(lubridate) #manipulating dates
```

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

```
## The following object is masked from 'package:base':
##
##     date
```

```
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:magrittr':
##
##     extract
```

```
library(pander)
```

```
file_path = "C:\\Users\\loren\\Desktop\\repos_github\\Covid19_prediction\\data\\comune_giorno.csv"
data1=read.csv(file_path)
#deleting 9999 missing values
data1=data1%>%subset(MASCHI_20 != 9999)
data1%>%head(20)
```

```
##    REG PROV NOME_REGIONE NOME_PROVINCIA NOME_COMUNE COD_PROVCOM
## 1    1    1     Piemonte         Torino       Agliè        1001
## 2    1    1     Piemonte         Torino       Agliè        1001
## 3    1    1     Piemonte         Torino       Agliè        1001
## 4    1    1     Piemonte         Torino       Agliè        1001
## 5    1    1     Piemonte         Torino       Agliè        1001
## 6    1    1     Piemonte         Torino       Agliè        1001
## 7    1    1     Piemonte         Torino       Agliè        1001
## 8    1    1     Piemonte         Torino       Agliè        1001
## 9    1    1     Piemonte         Torino       Agliè        1001
## 10   1    1     Piemonte         Torino       Agliè        1001
```

```
## 11   1    1     Piemonte         Torino       Agliè        1001
## 12   1    1     Piemonte         Torino       Agliè        1001
## 13   1    1     Piemonte         Torino       Agliè        1001
## 14   1    1     Piemonte         Torino       Agliè        1001
## 15   1    1     Piemonte         Torino       Agliè        1001
## 16   1    1     Piemonte         Torino       Agliè        1001
## 17   1    1     Piemonte         Torino       Agliè        1001
## 18   1    1     Piemonte         Torino       Agliè        1001
## 19   1    1     Piemonte         Torino       Agliè        1001
## 20   1    1     Piemonte         Torino       Agliè        1001
##    DATA_INIZIO_DIFF CL_ETA  GE MASCHI_15 MASCHI_16 MASCHI_17 MASCHI_18
## 1          1 aprile     17 102         0         0         0         0
## 2          1 aprile     18 104         0         0         0         0
## 3          1 aprile     18 105         0         0         0         0
## 4          1 aprile     17 106         1         0         0         0
## 5          1 aprile     18 106         0         0         0         1
## 6          1 aprile     20 106         0         0         0         0
## 7          1 aprile     16 107         0         0         0         0
## 8          1 aprile     17 107         0         0         0         0
## 9          1 aprile     21 107         0         0         0         0
## 10         1 aprile     19 108         0         0         0         0
## 11         1 aprile     19 109         0         0         0         0
## 12         1 aprile     20 109         0         0         0         0
## 13         1 aprile     14 110         0         0         0         0
## 14         1 aprile     10 113         0         0         0         0
## 15         1 aprile     17 117         0         0         0         0
## 16         1 aprile     19 117         0         0         0         0
## 17         1 aprile     19 118         0         0         0         0
## 18         1 aprile     20 118         0         0         0         0
## 19         1 aprile     19 119         0         1         0         0
## 20         1 aprile     13 120         0         0         0         0
##    MASCHI_19 MASCHI_20 FEMMINE_15 FEMMINE_16 FEMMINE_17 FEMMINE_18
## 1          0         0          0          0          0          1
## 2          0         0          0          1          0          0
## 3          0         1          0          0          0          0
## 4          0         0          0          0          0          0
## 5          0         0          0          0          0          0
## 6          0         0          0          0          0          1
## 7          0         0          0          0          0          0
## 8          0         0          0          0          1          0
## 9          0         0          0          1          0          0
## 10         0         0          0          1          0          0
## 11         0         0          0          0          0          1
## 12         0         0          0          0          0          0
## 13         0         1          0          0          0          0
## 14         1         0          0          0          0          0
## 15         0         0          0          1          0          0
## 16         0         0          0          0          0          1
## 17         0         0          0          0          0          0
## 18         0         0          1          0          0          0
## 19         0         0          0          0          0          0
## 20         0         0          0          1          0          0
##    FEMMINE_19 FEMMINE_20 TOTALE_15 TOTALE_16 TOTALE_17 TOTALE_18 TOTALE_19
## 1           0          0         0         0         0         1         0
```

```
## 2           0           0         0         1         0         0         0
## 3           0           0         0         0         0         0         0
## 4           0           0         1         0         0         0         0
## 5           0           0         0         0         0         1         0
## 6           0           0         0         0         0         1         0
## 7           1           0         0         0         0         0         1
## 8           0           0         0         0         1         0         0
## 9           0           0         0         1         0         0         0
## 10          0           0         0         1         0         0         0
## 11          0           0         0         0         0         1         0
## 12          0           1         0         0         0         0         0
## 13          0           0         0         0         0         0         0
## 14          0           0         0         0         0         0         1
## 15          0           0         0         1         0         0         0
## 16          0           0         0         0         0         1         0
## 17          0           1         0         0         0         0         0
## 18          0           0         1         0         0         0         0
## 19          0           0         0         1         0         0         0
## 20          0           0         0         1         0         0         0
##    TOTALE_20
## 1          0
## 2          0
## 3          1
## 4          0
## 5          0
## 6          0
## 7          0
## 8          0
## 9          0
## 10         0
## 11         0
## 12         1
## 13         1
## 14         0
## 15         0
## 16         0
## 17         1
## 18         0
## 19         0
## 20         0
```

```
#reshaping dataset
data2=gather(data1,'MASCHI_15','MASCHI_16','MASCHI_17','MASCHI_18','MASCHI_19','MASCHI_20','FEMMINE_15'

#elimino le righe con 0 morti e le colonne dei totali
data2=data2%>%subset(MORTI>0)
```

```
data3=data2%>%separate(`GENERE/ANNO`,into=c('GENERE','ANNO'),sep='_')

data3$GE=ifelse(data3$GE%>%nchar()==3,paste('0',data3$GE,data3$ANNO,sep=''),paste(data3$GE,data3$ANNO,s
data3$DATA=data3$GE%>%mdy()

data4=data3%>%select('DATA','NOME_REGIONE','NOME_PROVINCIA','NOME_COMUNE','GENERE','MORTI','CL_ETA','DA
```

```r
data4=data4%>%arrange(DATA)
```

```r
#excluding Apr 2020
data5= data4%>%subset(DATA%>%month() != 4)
data5$DATA%>%summary()%>%min()
```

```
## [1] "2015-01-01"
```

```r
data5$DATA%>%summary()%>%max()
```

```
## [1] "2020-03-31"
```

```r
cod=c(0:21)
tcod = c("0","1-4","5-9","10-14","15-19","20-24","25-29","30-34","35-39","40-44","45-49","50-54","55-59

for (i in 1:nrow(data5)){
  index = as.numeric(data5$CL_ETA[i])+1
  data5$CL_ETA[i]=tcod[index]
}

data_lomb = data5%>%filter(NOME_REGIONE=="Lombardia")

data_lomb%>%head(10)
```

```
##          DATA NOME_REGIONE NOME_PROVINCIA     NOME_COMUNE GENERE MORTI
## 1  2015-01-01    Lombardia         Varese         Cairate MASCHI     1
## 2  2015-01-01    Lombardia         Varese Cassano Magnago MASCHI     2
## 3  2015-01-01    Lombardia         Varese         Saronno MASCHI     1
## 4  2015-01-01    Lombardia         Varese          Varese MASCHI     1
## 5  2015-01-01    Lombardia           Como            Como MASCHI     1
## 6  2015-01-01    Lombardia           Como            Como MASCHI     1
## 7  2015-01-01    Lombardia           Como            Como MASCHI     1
## 8  2015-01-01    Lombardia        Sondrio          Chiuro MASCHI     1
## 9  2015-01-01    Lombardia        Sondrio        Morbegno MASCHI     1
## 10 2015-01-01    Lombardia        Sondrio         Sondalo MASCHI     1
##    CL_ETA DATA_INIZIO_DIFF COD_PROVCOM PROV REG     GE
## 1   85-89         1 aprile       12029   12   3 010115
## 2   75-79         8 aprile       12040   12   3 010115
## 3   90-94        16 aprile       12119   12   3 010115
## 4   90-94         8 aprile       12133   12   3 010115
## 5   65-69         8 aprile       13075   13   3 010115
## 6   75-79         8 aprile       13075   13   3 010115
## 7   90-94         8 aprile       13075   13   3 010115
## 8   80-84         8 aprile       14020   14   3 010115
## 9   75-79         1 aprile       14045   14   3 010115
## 10  75-79         1 aprile       14060   14   3 010115
```

```r
data_lomb$GENERE=ifelse(data_lomb$GENERE=="MASCHI","M","F")
data_lomb%>%head(10)
```

```
##          DATA NOME_REGIONE NOME_PROVINCIA     NOME_COMUNE GENERE MORTI
## 1  2015-01-01    Lombardia         Varese         Cairate      M     1
## 2  2015-01-01    Lombardia         Varese Cassano Magnago      M     2
## 3  2015-01-01    Lombardia         Varese         Saronno      M     1
## 4  2015-01-01    Lombardia         Varese          Varese      M     1
## 5  2015-01-01    Lombardia           Como            Como      M     1
```

```
## 6  2015-01-01    Lombardia           Como            Como      M     1
## 7  2015-01-01    Lombardia           Como            Como      M     1
## 8  2015-01-01    Lombardia        Sondrio          Chiuro      M     1
## 9  2015-01-01    Lombardia        Sondrio        Morbegno      M     1
## 10 2015-01-01    Lombardia        Sondrio         Sondalo      M     1
##    CL_ETA DATA_INIZIO_DIFF COD_PROVCOM PROV REG     GE
## 1   85-89         1 aprile       12029   12   3 010115
## 2   75-79         8 aprile       12040   12   3 010115
## 3   90-94        16 aprile       12119   12   3 010115
## 4   90-94         8 aprile       12133   12   3 010115
## 5   65-69         8 aprile       13075   13   3 010115
## 6   75-79         8 aprile       13075   13   3 010115
## 7   90-94         8 aprile       13075   13   3 010115
## 8   80-84         8 aprile       14020   14   3 010115
## 9   75-79         1 aprile       14045   14   3 010115
## 10  75-79         1 aprile       14060   14   3 010115
```