Descriptive_Statistics

OMR Group 2

30-11-2020

```
library(ISLR)
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(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v stringr 1.4.0
## v tidyr   1.1.1     v forcats 0.5.0
## v readr   1.4.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(haven)
library(readxl)
library(Matrix)
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
library(tinytex)
library(ggplot2)
library(wesanderson)
library(png)
```

Import data and check data format.

```
data <- read.csv(unz("../data/30-11-2020.zip", "30-11-2020.csv"), header = TRUE)

Contact_matrix <- read.csv("../data/Contact_matrix.csv", header = TRUE)

class <- lapply(data,class)
```

```
##Build contact matrix for the physical&non-physical contact of people in different ag

Contact_matrix <- Contact_matrix %>%
                    rename("0_4" = 1, "5_9" = 2, "10_14" = 3, "15_19" = 4, "20_24" = 5,
```

Begin cleaning the data.

```
#Create data frame for infection rate, hence '_i'
#Mutate 'Agegroup' to fix the formatting issue
#Mutate a further two times to plot values in order later
data_i <- data %>%
  filter(Agegroup != "Unknown") %>%
  filter(Agegroup != "<50") %>%
  filter(Sex != "Unknown") %>%
  mutate(Agegroup=recode(Agegroup, `Oct-19` = '10-19')) %>%
  mutate(Agegroup = as.character(Agegroup)) %>%
  mutate(Agegroup = as.factor(Agegroup))

#Create data frame with infections of last 10 days
data_i10 <- data %>%
  filter(Date_statistics == c("31/10/2020", "30/10/2020", "29/10/2020", "28/10/2020", "2
  filter(Agegroup != "Unknown") %>%
  filter(Agegroup != "<50") %>%
  filter(Sex != "Unknown") %>%
  mutate(Agegroup=recode(Agegroup, `Oct-19` = '10-19')) %>%
  mutate(Agegroup = as.character(Agegroup)) %>%
  mutate(Agegroup = as.factor(Agegroup))
```

```
## Warning in `==.default`(Date_statistics, c("31/10/2020", "30/10/2020",
## "29/10/2020", : longer object length is not a multiple of shorter object length
```

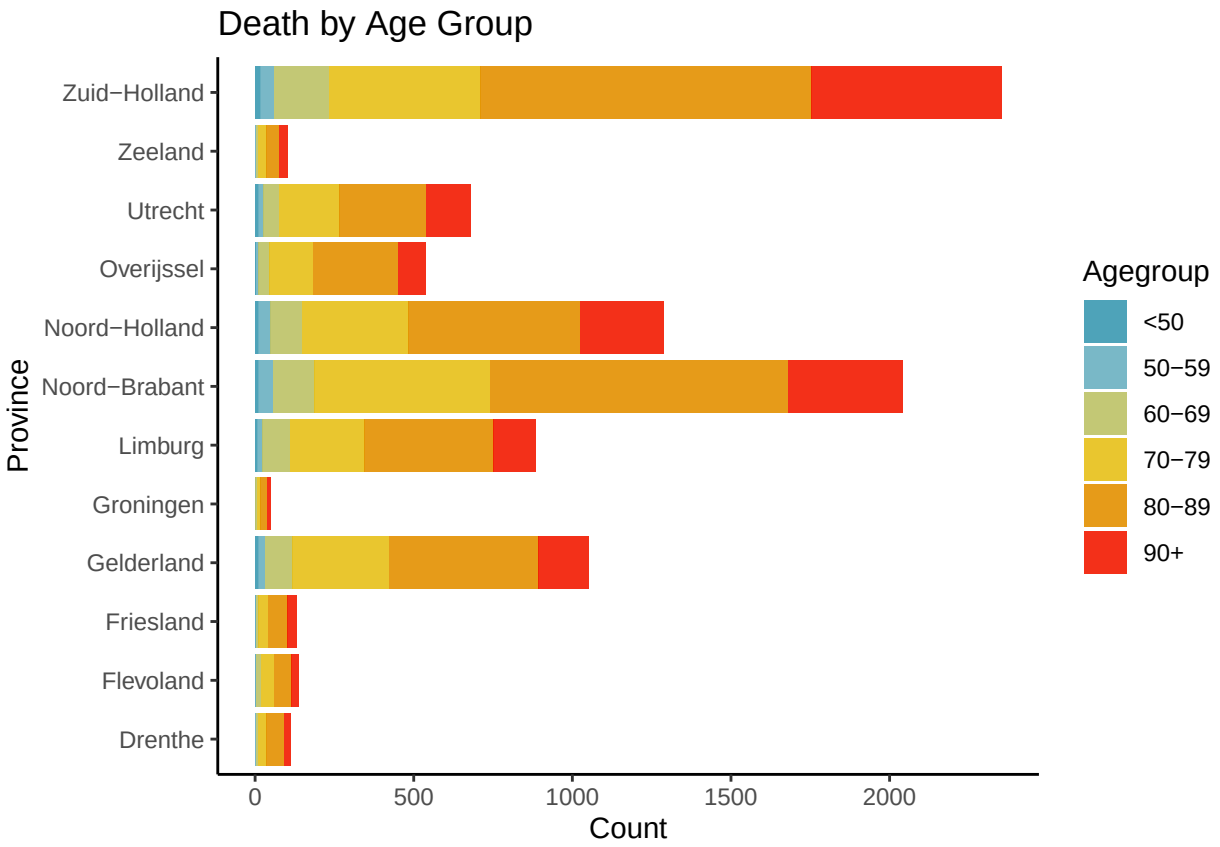
```
## Warning in is.na(e1) | is.na(e2): longer object length is not a multiple of
## shorter object length
```

```r
#Create data frame for death rate, hence '_d'
data_d <- data %>%
  filter(Agegroup != "Unknown") %>%
  filter(Sex != "Unknown") %>%
  filter(!is.na(Week_of_death))
```

```r
#Plot the number of infections using 'data_i' data frame
infections_plot <- data_i %>%
  arrange(Agegroup) %>%
  group_by(Province) %>%
  ggplot(aes(x=Province)) +
  #guides(col=FALSE) +
  geom_bar(mapping = aes(fill = Agegroup), alpha=0.9, position = position_stack(reverse
  scale_fill_manual(values = wes_palette("Zissou1", 10, type = "continuous")) +
  coord_flip() +
  theme_classic() +
  labs(x="Province", y="Count",
      title="Infection by Age Group")

#Save infections rate plot
#png(infections_plot = "/images/Contact_matrix.png")
#plot(infections_plot)
#dev.off()
```

```r
#Plot the numbe of deaths using the 'data_d' data frame
data_d %>%
  group_by(Province) %>%
  ggplot(aes(x=Province)) +
  #guides(col=FALSE) +
  geom_bar(mapping = aes(fill = Agegroup), alpha=0.9, position = position_stack(reverse
  scale_fill_manual(values = wes_palette("Zissou1", 6, type = "continuous")) +
  coord_flip() +
  theme_classic() +
  labs(x="Province", y="Count",
      title="Death by Age Group")
```




```
##Create matrix for proportion of infections by age group
#Extract age groups column from infections data frame
agg_i <- aggregate(data_i$Agegroup, by=list(data_i$Agegroup), FUN=length)

agg_i <- agg_i %>%
          rename("Count" = "x") %>%
          rename("Agegroup" = "Group.1") %>%
    group_by(Agegroup) %>%
    mutate(Sum = sum(agg_i$x)) %>%
    mutate(Proportion=(round(Count/Sum,4)))

#Diagonal matrix
diag_i <- Diagonal(nrow(agg_i), agg_i$Count)

#Diagonal matrix for proportion of infections
diag_i_prop <- Diagonal(nrow(agg_i), agg_i$Proportion)

##Create matrix for proportion of deaths by age group
#Extract age groups column from death rate data frame
agg_d <- aggregate(data_d$Agegroup, by=list(data_d$Agegroup), FUN=length)
```

```r
agg_d <- agg_d %>%
          rename("Count" = "x") %>%
          rename("Agegroup" = "Group.1") %>%
    group_by(Agegroup) %>%
    mutate(Sum = sum(agg_d$x)) %>%
    mutate(Proportion=(round(Count/Sum,4)))

#Diagonal matrix
diag_d <- Diagonal(nrow(agg_d), agg_d$Count)

#Diagonal matrix for proportion of deaths
diag_d_prop <- Diagonal(nrow(agg_d), agg_d$Proportion)
```

```r
agg_i10 <- aggregate(data_i10$Agegroup, by=list(data_i10$Agegroup), FUN=length)

agg_i10 <- agg_i10 %>%
          rename("Count" = "x") %>%
          rename("Agegroup" = "Group.1") %>%
    group_by(Agegroup) %>%
    mutate(Sum = sum(agg_i10$x)) %>%
    mutate(Proportion=(round(Count/Sum,4)))

#agg_i10 = I_zero
#Now construct R_0 from agg_i and agg_i10
R0 = (agg_i$Count - agg_i10$Count)

#Now scale by randomized testing
#Extract data from papers
```