assignment_1_Rmd

Emil Hafeez (eh2928)

9/18/2020

#Assignment 1 Assignment 1 for Biostatistics P8130, using a dataset callled Antibodies, practices data manipulation and descriptive statistics (including table, histogram, and boxplot generation). The dataset contains information on demographic variables, IgM antibodies, and self-reported smell loss for patients diagnosed with COVID-19 (via the PCR gold-standard).

Let's load the appropriate packages first.

```
library(arsenal)
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(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v tibble  3.0.3     v purrr   0.3.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

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

##Question 1

```
#Read the CSV data into a dataframe
antibodies_df <- read.csv("/Users/emilhafeez/Google Drive/Columbia/Fall 2020/Classes/Biostatistics 1/Ass

#Examine basics of the data, to get to know it
names(antibodies_df)
```

```
## [1] "Subject"      "AgeCategory"  "Antibody_IgM" "Smell"        "Gender"
```

```
nrow(antibodies_df)
```

```
## [1] 1491
```

```r
ncol(antibodies_df)
```

```
## [1] 5
```

```r
head(antibodies_df)
```

```
##   Subject AgeCategory Antibody_IgM  Smell Gender
## 1       1       31-50           NA Normal Female
## 2       2       18-30           NA Normal Female
## 3       3       18-30           NA Normal   Male
## 4       4       31-50           NA Normal   Male
## 5       5       31-50           NA Normal   Male
## 6       6       31-50        0.055 Normal   Male
```

```r
tail(antibodies_df)
```

```
##      Subject AgeCategory Antibody_IgM   Smell Gender
## 1486    2911       18-30           NA Altered Female
## 1487    2912         51+           NA Altered   Male
## 1488    2913       18-30           NA Altered Female
## 1489    2914       31-50           NA Altered Female
## 1490    2915       31-50           NA Altered Female
## 1491    2917       31-50           NA  Normal Female
```

```r
anyNA(antibodies_df)
```

```
## [1] TRUE
```

```r
str(antibodies_df)
```

```
## 'data.frame':    1491 obs. of  5 variables:
##  $ Subject     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ AgeCategory : chr  "31-50" "18-30" "18-30" "31-50" ...
##  $ Antibody_IgM: num  NA NA NA NA NA 0.055 NA NA NA NA ...
##  $ Smell       : chr  "Normal" "Normal" "Normal" "Normal" ...
##  $ Gender      : chr  "Female" "Female" "Male" "Male" ...
```