Lesson 6 Homework

Leon Woltermann

```
library(haven)
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)
```

```
file <- read_dta('Bosker_Data/bagdad-london--finalRestat.dta')
file
```

```
## # A tibble: 8,723 x 69
##    indicator city   country  year arab_peninsula latitude longitude citypop_le10
##        <dbl> <chr>  <chr>   <dbl>          <dbl>    <dbl>     <dbl>        <dbl>
##  1         1 Graz   Austria   800              0     47.1      15.4            0
##  2         2 Innsb~ Austria   800              0     47.3      11.4            0
##  3         3 Klage~ Austria   800              0     46.6      14.3            0
##  4         4 Linz   Austria   800              0     48.3      14.3            0
##  5         5 Salzb~ Austria   800              0     47.8      13.0            0
##  6         6 Wien ~ Austria   800              0     48.2      16.4            0
##  7         7 Aalst~ Belgium   800              0     50.9       4.03           0
##  8         8 Antwe~ Belgium   800              0     51.2       4.42           0
##  9         9 Brugg~ Belgium   800              0     51.2       3.23           0
## 10        10 Bruxe~ Belgium   800              0     50.8       4.33           0
## # ... with 8,713 more rows, and 61 more variables: citypop_le5 <dbl>,
## #   sea <dbl>, river <dbl>, hub_3rr <dbl>, rom_road_nohub <dbl>,
## #   caravan_hub <dbl>, caravan_nohub <dbl>, elevation_m <dbl>, rugg10 <dbl>,
## #   bishop <dbl>, archbishop <dbl>, capital <dbl>, university <dbl>,
## #   muslim <dbl>, me_na <dbl>, muslim_holy_city <dbl>, plundered <dbl>,
## #   soilquality <dbl>, commune <dbl>, ecozones <dbl>, free_prince_dls <dbl>,
## #   total_pop_country <dbl>, dmedina <dbl>, dmecca <dbl>, drome <dbl>, ...
```

```r
#What is the chronological extent of this data?
min(file$year)
```

```
## [1] 800
```

```r
max(file$year)
```

```
## [1] 1800
```

```r
#or

summary(file$year)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     800    1000    1300    1300    1600    1800
```

```r
#What periods can it be divided into? How can we do that?

#First approach:
#Create a vector with all the periods in the data set.
#Create a vector with the labels for the periods according to the rows of the first vector.
#Then create a dataframe from both vectors and left join it with the main data set.

year <- c(seq(800, 1800, 100))
period <- c("Middle Ages",  "Middle Ages", "Middle Ages", "Middle Ages", "Middle Ages",
            "Middle Ages", "Middle Ages", "Modern Times", "Modern Times",
            "Modern Times", "Modern Times")

periods <- data.frame(year, period)

file2 <- file %>%
  left_join(periods, by="year")
head(file2)
```

```
## # A tibble: 6 x 70
##   indicator city    country  year arab_peninsula latitude longitude citypop_le10
##       <dbl> <chr>   <chr>   <dbl>          <dbl>    <dbl>     <dbl>        <dbl>
## 1         1 Graz    Austria   800              0     47.1      15.4            0
## 2         2 Innsbr~ Austria   800              0     47.3      11.4            0
## 3         3 Klagen~ Austria   800              0     46.6      14.3            0
## 4         4 Linz    Austria   800              0     48.3      14.3            0
## 5         5 Salzbu~ Austria   800              0     47.8      13.0            0
## 6         6 Wien (~ Austria   800              0     48.2      16.4            0
## # ... with 62 more variables: citypop_le5 <dbl>, sea <dbl>, river <dbl>,
## #   hub_3rr <dbl>, rom_road_nohub <dbl>, caravan_hub <dbl>,
## #   caravan_nohub <dbl>, elevation_m <dbl>, rugg10 <dbl>, bishop <dbl>,
## #   archbishop <dbl>, capital <dbl>, university <dbl>, muslim <dbl>,
## #   me_na <dbl>, muslim_holy_city <dbl>, plundered <dbl>, soilquality <dbl>,
## #   commune <dbl>, ecozones <dbl>, free_prince_dls <dbl>,
## #   total_pop_country <dbl>, dmedina <dbl>, dmecca <dbl>, drome <dbl>, ...
```

```r
#What periods can it be divided into? How can we do that?

#Second approach:
#Create a new column "century" with mutate based on conditions.


library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v tibble  3.1.6     v purrr   0.3.4
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

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

```r
year2 <- c(seq(800, 1800, 1))

centuries <- tibble(year2)

centuries %>%
  mutate(century = ifelse(year2<800, "8th Century",
                          ifelse(year2<900,"9th Century",
                                 ifelse(year2<1000, "10th Century",
                                        ifelse(year2<1100,"11th Century",
                                               ifelse(year2<1200,"12th Century",
                                                      ifelse(year2<1300,"13th Century",
                                                             ifelse(year2<1400,"14th Century",
                                                                    ifelse(year2<1500,"15th Century",
                                                                           ifelse(year2<1600,"16th Centu
                                                                                  ifelse(year2<1700,"17
                                                                                         ifelse(year2<18
```

```
## # A tibble: 1,001 x 2
##    year2 century
##    <dbl> <chr>
##  1   800 9th Century
##  2   801 9th Century
##  3   802 9th Century
##  4   803 9th Century
##  5   804 9th Century
##  6   805 9th Century
##  7   806 9th Century
##  8   807 9th Century
##  9   808 9th Century
## 10   809 9th Century
## # ... with 991 more rows
```

```
head(centuries)
```

```
## # A tibble: 6 x 1
##   year2
##   <dbl>
## 1   800
## 2   801
## 3   802
## 4   803
## 5   804
## 6   805
```

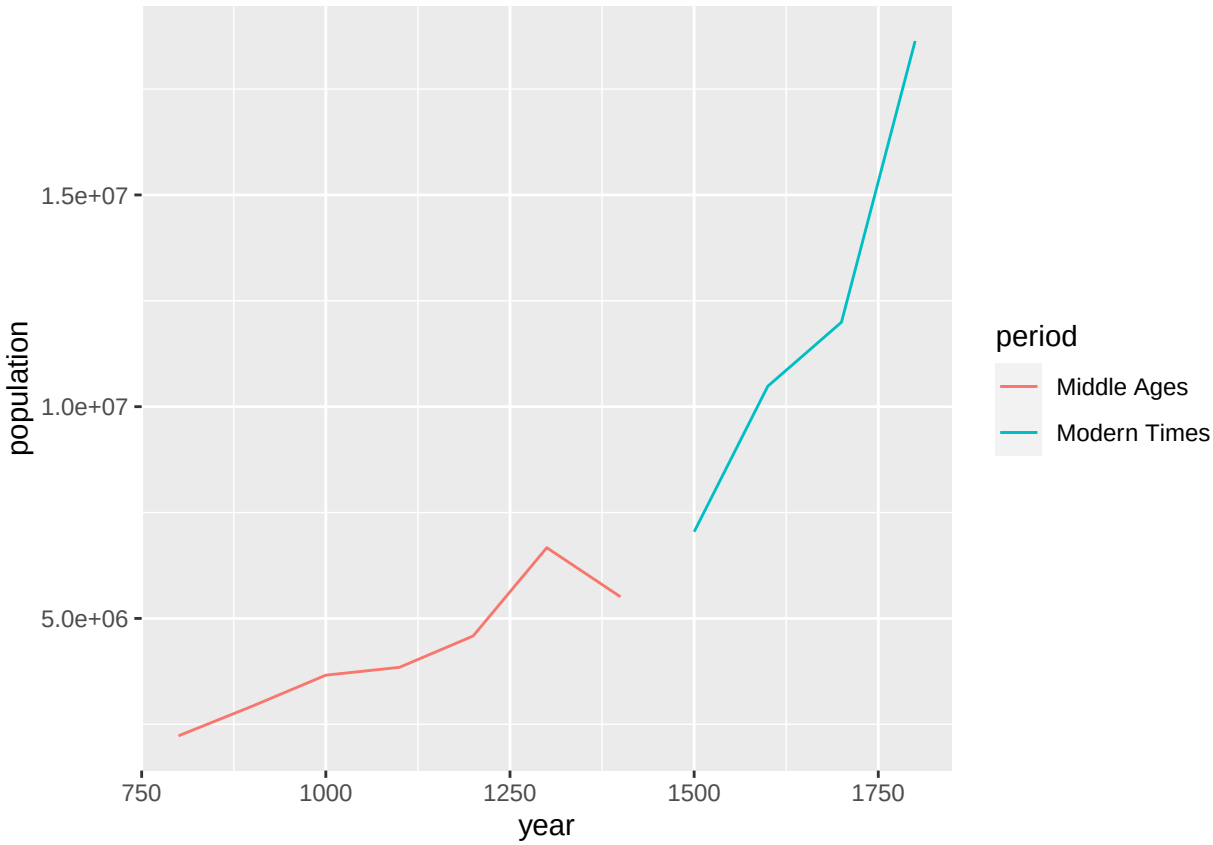
```
#Can you generate a cumulative graph of population over time, divided into these periods?

#Summarize population grouped by year and period.
population_growth <- file2 %>%
  select(city, country, year, citypop_le10, citypop_le5, period) %>%
  group_by(year, period) %>%
  summarize(population = sum(citypop_le5 * 1000))
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
population_growth
```

```
## # A tibble: 11 x 3
## # Groups:   year [11]
##     year period       population
##    <dbl> <chr>             <dbl>
##  1   800 Middle Ages     2227000
##  2   900 Middle Ages     2926000
##  3  1000 Middle Ages     3659000
##  4  1100 Middle Ages     3844000
##  5  1200 Middle Ages     4584000
##  6  1300 Middle Ages     6670000
##  7  1400 Middle Ages     5510000
##  8  1500 Modern Times    7046000
##  9  1600 Modern Times   10483000
## 10  1700 Modern Times   11994000
## 11  1800 Modern Times   18635000
```

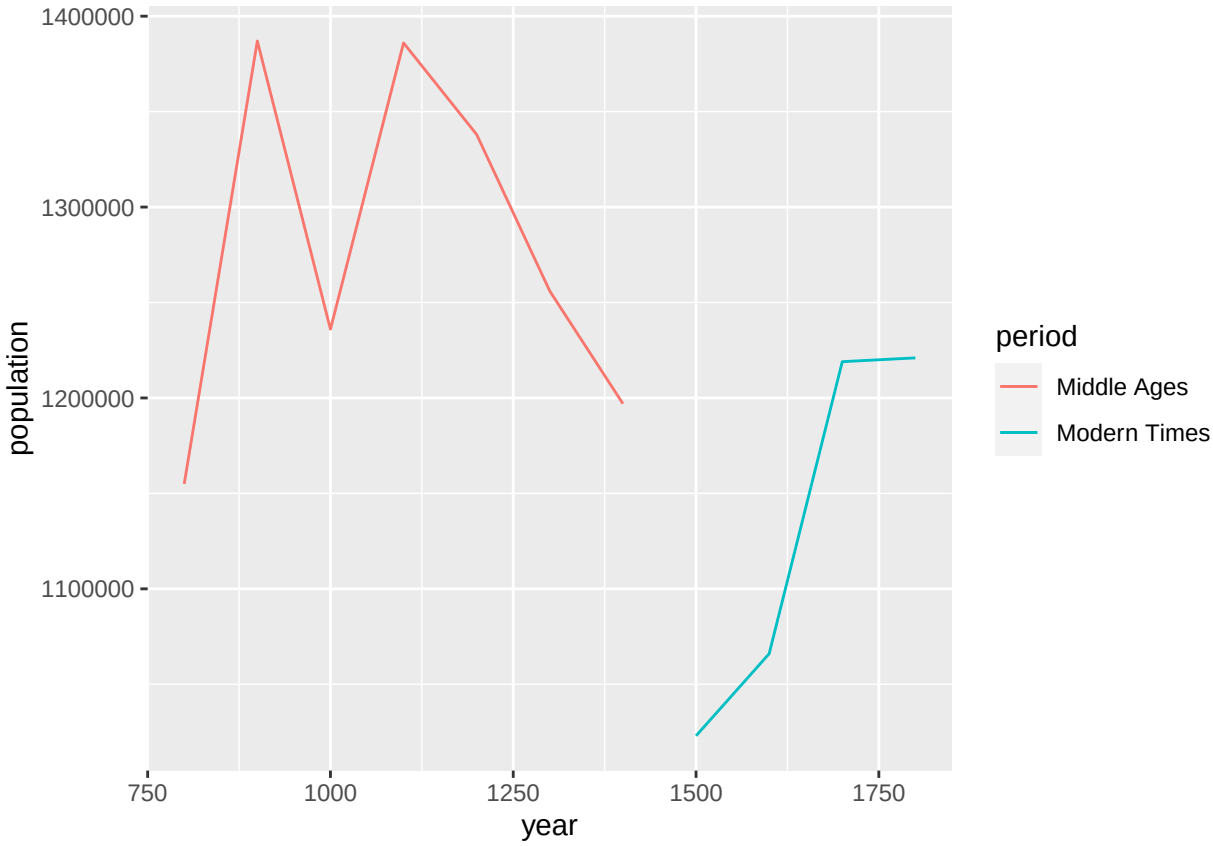
```
#Select the periods, so they can be plotted separately.
middle_ages_population <- population_growth %>%
  filter(period == "Middle Ages")
modern_times_population <- population_growth %>%
  filter(period == "Modern Times")


ggplot() + geom_line(data=middle_ages_population, aes(x= year, y=population, color=period)) + geom_line
```




```
#population in north african cities

#the variable me_na indicates whether or not a city is located in the Middle East or North Africa
#the variable arab_peninsula indicates whether or not a city is located on the Arabian Peninsula
#turkey is filtered out manually
northAfrica_citites <- file2 %>%
  filter(me_na == 1) %>%
  filter(arab_peninsula != 1) %>%
  filter(country != "Turkey")

#Summarize population grouped by year and period.
population_growth_northAfrica <- northAfrica_citites %>%
  select(city, country, year, citypop_le10, citypop_le5, period) %>%
  group_by(year, period) %>%
  summarize(population = sum(citypop_le5 * 1000))
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

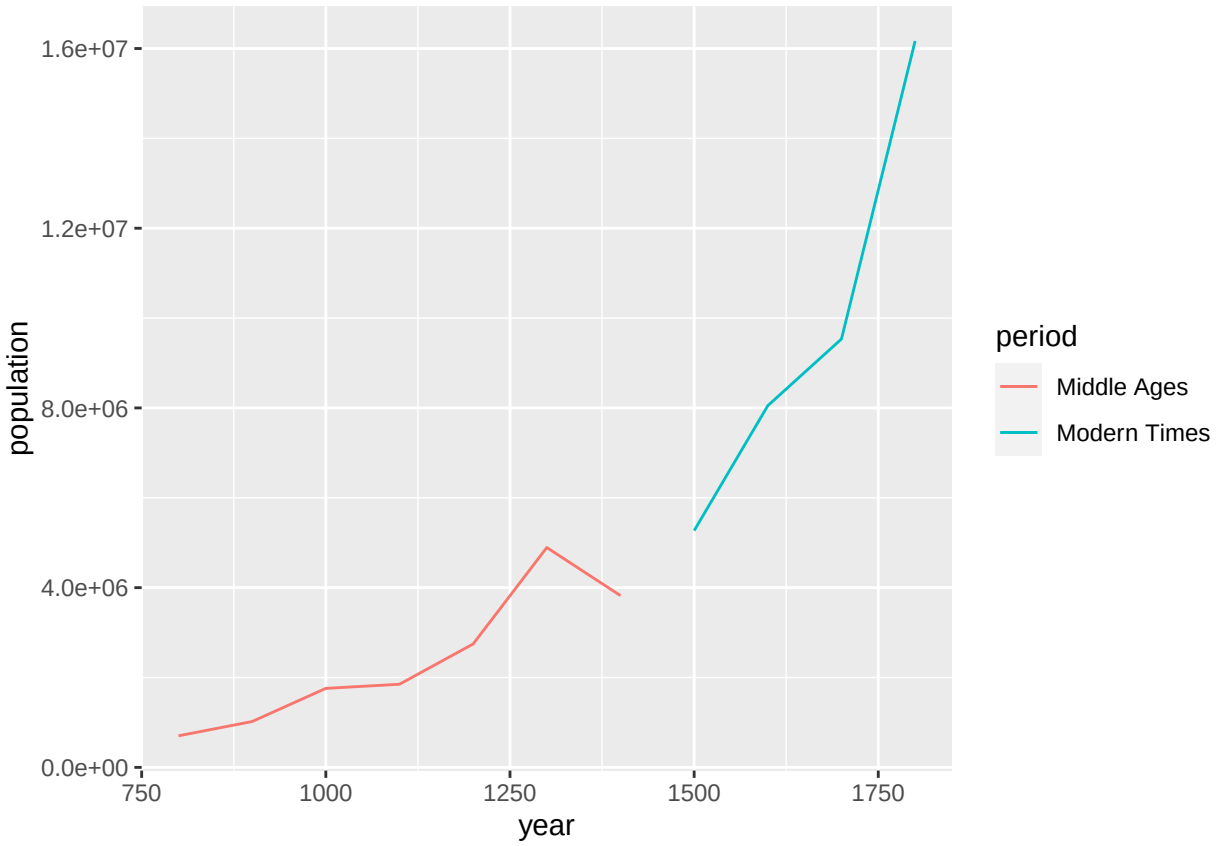
```
#Select the periods, so they can be plotted separately.
middle_ages_northAfrica <- population_growth_northAfrica %>%
  filter(period == "Middle Ages")
modern_times_northAfrica <- population_growth_northAfrica %>%
  filter(period == "Modern Times")

ggplot() + geom_line(data=middle_ages_northAfrica, aes(x= year, y=population, color=period)) + geom_lin
```




```
#Population in European cities

#I don't know if it is that simple.
#Since me_na indicates whether or not a city is located in the Middle East or North Africa,
#disabling it would show any other city except in this region.
european_cities <- file2 %>%
  filter(me_na == 0)

#After checking it with unique(european_cities$country) it seems to be correct.

population_growth_europe <- european_cities %>%
  select(city, country, year, citypop_le10, citypop_le5, period) %>%
  group_by(year, period) %>%
  summarize(population = sum(citypop_le5 * 1000))
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
middle_ages_europe <- population_growth_europe %>%
  filter(period == "Middle Ages")
modern_times_europe <- population_growth_europe %>%
  filter(period == "Modern Times")

ggplot() + geom_line(data=middle_ages_europe, aes(x= year, y=population, color=period)) + geom_line(data
```




```
#Ottoman empire
ottoman_citites <- file2 %>%
  filter(ottoman == 1)

#We want to find out when there were the most cities in a century or in other words count the observati
number_cities_perYear <- ottoman_citites %>%
  group_by(year) %>%
  summarize(citites_per_century = n())

#largest number of cities
#The which.max() function returns the maximum value of a column.
number_cities_perYear[which.max(number_cities_perYear$citites_per_century),]
```

```
## # A tibble: 1 x 2
##    year citites_per_century
##   <dbl>               <int>
## 1  1600                 160
```

```
#When was its population at the highest?
population_ottoman <- ottoman_citites %>%
  select(city, country, year, citypop_le5) %>%
  group_by(year) %>%
  summarize(population = sum(citypop_le5 * 1000))

#population_ottoman
population_ottoman[which.max(population_ottoman$population),]
```

```
## # A tibble: 1 x 2
##    year population
##   <dbl>      <dbl>
## 1  1800    2873000
```

```
#Christiandom and Islamdom?

#The variable muslim indicates whether or not a city falls under Muslim rule.
#I am not sure whether this is sufficient in this dataset,
#as Islamdom signifies the places (cities) where Islam is
#the predominant religion and not under Muslim rule.
#Vice versa Muslim rule does not indicate that Islam is the predominant religion.
islam_citites <- file2 %>%
  filter(muslim == 1)

#Here as well, I am not sure whether disabling the muslim variable gives us Christiandom
christian_citites <- file2 %>%
   filter(muslim == 0)

#What are the largest cities of Islamdom for each reported period?
#the filter returns the row with the biggest value in citypop_le10 of each group
islam_citites %>%
  group_by(year) %>%
  filter(citypop_le10==max(citypop_le10))
```

```
## # A tibble: 12 x 70
## # Groups:   year [11]
##    indicator city   country  year arab_peninsula latitude longitude citypop_le10
##        <dbl> <chr>  <chr>   <dbl>          <dbl>    <dbl>     <dbl>        <dbl>
##  1       619 Baghd~ Iraq      800              0     33.3      44.4          350
##  2       619 Baghd~ Iraq      900              0     33.3      44.4          450
##  3       619 Baghd~ Iraq     1000              0     33.3      44.4          300
##  4       619 Baghd~ Iraq     1100              0     33.3      44.4          250
##  5       619 Baghd~ Iraq     1200              0     33.3      44.4          200
##  6       624 Fusta~ Egypt    1200              0     30.0      31.2          200
##  7       624 Fusta~ Egypt    1300              0     30.0      31.2          220
##  8       624 Fusta~ Egypt    1400              0     30.0      31.2          250
##  9       598 Const~ Turkey   1500              0     41.0      29.0          280
## 10       598 Const~ Turkey   1600              0     41.0      29.0          700
## 11       598 Const~ Turkey   1700              0     41.0      29.0          700
## 12       598 Const~ Turkey   1800              0     41.0      29.0          500
## # ... with 62 more variables: citypop_le5 <dbl>, sea <dbl>, river <dbl>,
## #   hub_3rr <dbl>, rom_road_nohub <dbl>, caravan_hub <dbl>,
## #   caravan_nohub <dbl>, elevation_m <dbl>, rugg10 <dbl>, bishop <dbl>,
## #   archbishop <dbl>, capital <dbl>, university <dbl>, muslim <dbl>,
## #   me_na <dbl>, muslim_holy_city <dbl>, plundered <dbl>, soilquality <dbl>,
## #   commune <dbl>, ecozones <dbl>, free_prince_dls <dbl>,
## #   total_pop_country <dbl>, dmedina <dbl>, dmecca <dbl>, drome <dbl>, ...
```

```
#What are the largest western cities of Islamdom between 1000 and 1500 CE?
islam_citites %>%
  filter(me_na != 1) %>%
  group_by(year) %>%
  filter(citypop_le10==max(citypop_le10))
```

```
## # A tibble: 12 x 70
## # Groups:   year [11]
##    indicator city   country  year arab_peninsula latitude longitude citypop_le10
##        <dbl> <chr>  <chr>   <dbl>          <dbl>    <dbl>     <dbl>         <dbl>
##  1       443 Cordo~ Spain     800              0     37.9     -4.77            75
##  2       443 Cordo~ Spain     900              0     37.9     -4.77            95
##  3       443 Cordo~ Spain    1000              0     37.9     -4.77           100
##  4       502 Sevil~ Spain    1100              0     37.4     -5.99            85
##  5       502 Sevil~ Spain    1200              0     37.4     -5.99            80
##  6       450 Grana~ Spain    1300              0     37.2     -3.60           150
##  7       450 Grana~ Spain    1400              0     37.2     -3.60           100
##  8       659 Skopj~ Yugosl~  1500              0     42.0     21.4             50
##  9       650 Belgr~ Yugosl~  1600              0     44.8     20.5             55
## 10       659 Skopj~ Yugosl~  1600              0     42.0     21.4             55
## 11       650 Belgr~ Yugosl~  1700              0     44.8     20.5             50
## 12       596 Salon~ Greece   1800              0     40.6     22.9             70
## # ... with 62 more variables: citypop_le5 <dbl>, sea <dbl>, river <dbl>,
## #   hub_3rr <dbl>, rom_road_nohub <dbl>, caravan_hub <dbl>,
## #   caravan_nohub <dbl>, elevation_m <dbl>, rugg10 <dbl>, bishop <dbl>,
## #   archbishop <dbl>, capital <dbl>, university <dbl>, muslim <dbl>,
## #   me_na <dbl>, muslim_holy_city <dbl>, plundered <dbl>, soilquality <dbl>,
## #   commune <dbl>, ecozones <dbl>, free_prince_dls <dbl>,
## #   total_pop_country <dbl>, dmedina <dbl>, dmecca <dbl>, drome <dbl>, ...
```