MidTerm2-Part2

Gabriel Ndoum

11/8/2020

## Description:

For this Question, use the file all-states-history.data.csv that contains covid-19 tracking data of covid-19 positive and death cases in The United States. This file is a simpler version of the data found in this website <https://covidtracking.com/data>

## Importing Data & libraries

library(purrr)

## Warning: package 'purrr' was built under R version 4.0.3

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

## -- Conflicts ------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

mydata <- read.csv('coviddata.csv') #read csv file  
  
head(mydata)

## Month Day Year state deathIncrease hospitalizedCurrently inIcuCurrently  
## 1 10 28 2020 AK 1 80 NA  
## 2 10 28 2020 AL 19 1013 NA  
## 3 10 28 2020 AR 18 644 228  
## 4 10 28 2020 AS 0 NA NA  
## 5 10 28 2020 AZ 14 871 188  
## 6 10 28 2020 CA 75 2998 770  
## onVentilatorCurrently positiveIncrease totalTestResultsIncrease  
## 1 8 357 3170  
## 2 NA 1269 7419  
## 3 94 961 9706  
## 4 NA 0 0  
## 5 94 1043 6800  
## 6 NA 4515 96547

## Total deaths in the United States in each month

mydata %>%  
 group\_by(Month) %>%  
 summarise(death = sum(deathIncrease)) %>%  
 arrange(Month)

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

## # A tibble: 10 x 2  
## Month death  
## <int> <int>  
## 1 1 0  
## 2 2 5  
## 3 3 4308  
## 4 4 55264  
## 5 5 40918  
## 6 6 19566  
## 7 7 25286  
## 8 8 30265  
## 9 9 23313  
## 10 10 20449

## Total positive cases in the United States in each month.

mydata %>%  
 group\_by(Month) %>%  
 summarise(cases = sum(positiveIncrease)) %>%  
 arrange(Month)

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

## # A tibble: 10 x 2  
## Month cases  
## <int> <int>  
## 1 1 0  
## 2 2 18  
## 3 3 197442  
## 4 4 876049  
## 5 5 719691  
## 6 6 836077  
## 7 7 1903331  
## 8 8 1460595  
## 9 9 1197024  
## 10 10 1611763

## Total tests performed in the United States in each month.

mydata %>%  
 group\_by(Month) %>%  
 summarise(tests = sum(totalTestResultsIncrease)) %>%  
 arrange(Month)

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

## # A tibble: 10 x 2  
## Month tests  
## <int> <int>  
## 1 1 6  
## 2 2 73  
## 3 3 1126621  
## 4 4 5350492  
## 5 5 11049873  
## 6 6 15926538  
## 7 7 24812979  
## 8 8 24571000  
## 9 9 26233218  
## 10 10 30434682

## Total number of deaths in the United States till date.

sum(mydata$deathIncrease)

## [1] 219374

## Compute total deaths in each month for Texas and New York then, visualize .

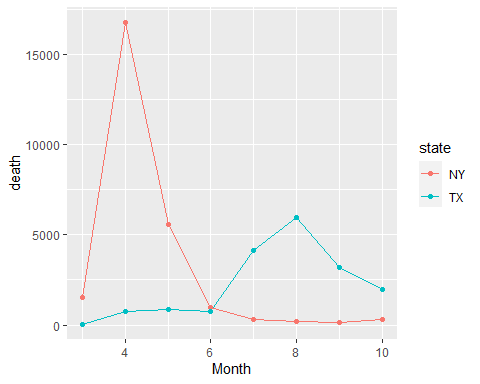
# Computing the total deaths in each month for the Texas AND New York states only  
TXNY\_death<- mydata %>%  
 group\_by(Month,state) %>%  
 summarize(death = sum(deathIncrease,na.rm = TRUE)) %>%  
 filter(state %in% c("TX","NY"))

## `summarise()` regrouping output by 'Month' (override with `.groups` argument)

print(TXNY\_death)

## # A tibble: 16 x 3  
## # Groups: Month [8]  
## Month state death  
## <int> <chr> <int>  
## 1 3 NY 1550  
## 2 3 TX 41  
## 3 4 NY 16771  
## 4 4 TX 741  
## 5 5 NY 5584  
## 6 5 TX 890  
## 7 6 NY 950  
## 8 6 TX 752  
## 9 7 NY 295  
## 10 7 TX 4145  
## 11 8 NY 178  
## 12 8 TX 5967  
## 13 9 NY 151  
## 14 9 TX 3175  
## 15 10 NY 294  
## 16 10 TX 1989

# Visualization  
ggplot(TXNY\_death) +  
 geom\_line(aes(x = Month, y = death, col = state)) +  
 geom\_point(aes(x = Month, y = death, col = state))



## Compute total positive cases in each month for Texas and New York then,

visualize

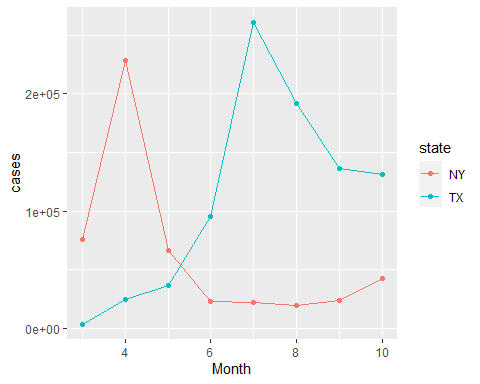
# Computing the total positive cases in each month for the Texas AND New York states only  
TXNY\_posCases<- mydata %>%  
 group\_by(Month,state) %>%  
 summarize(cases = sum(positiveIncrease,na.rm = TRUE)) %>%  
 filter(state %in% c("TX","NY"))

## `summarise()` regrouping output by 'Month' (override with `.groups` argument)

print(TXNY\_posCases)

## # A tibble: 16 x 3  
## # Groups: Month [8]  
## Month state cases  
## <int> <chr> <int>  
## 1 3 NY 75789  
## 2 3 TX 3265  
## 3 4 NY 228577  
## 4 4 TX 24821  
## 5 5 NY 66398  
## 6 5 TX 36200  
## 7 6 NY 22684  
## 8 6 TX 95699  
## 9 7 NY 21560  
## 10 7 TX 260960  
## 11 8 NY 19742  
## 12 8 TX 192023  
## 13 9 NY 23893  
## 14 9 TX 135998  
## 15 10 NY 42028  
## 16 10 TX 131027

# Visualization  
ggplot(TXNY\_posCases) +  
 geom\_line(aes(x = Month, y = cases, col = state)) +  
 geom\_point(aes(x = Month, y = cases, col = state))



## Compute positivity rate in each month (100\*positive cases in the month/total tests in the month) for Texas and New York. Then, visualize.

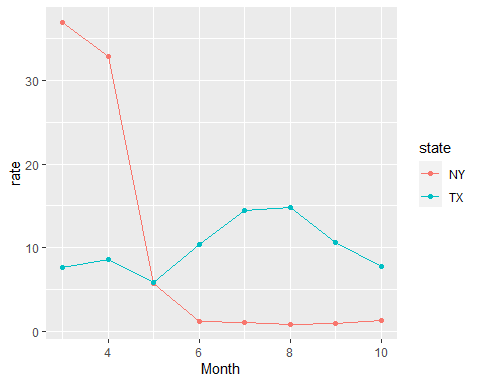
# Computing the total positive cases in each month for the Texas AND New York states only  
TXNY\_rate<- mydata %>%  
 group\_by(Month,state) %>%  
 summarize(rate = 100\*sum(positiveIncrease,na.rm = TRUE)/sum(totalTestResultsIncrease,na.rm = TRUE)) %>%  
 filter(state %in% c("TX","NY"))

## `summarise()` regrouping output by 'Month' (override with `.groups` argument)

print(TXNY\_rate)

## # A tibble: 16 x 3  
## # Groups: Month [8]  
## Month state rate  
## <int> <chr> <dbl>  
## 1 3 NY 36.9   
## 2 3 TX 7.59   
## 3 4 NY 32.9   
## 4 4 TX 8.64   
## 5 5 NY 5.71   
## 6 5 TX 5.82   
## 7 6 NY 1.23   
## 8 6 TX 10.4   
## 9 7 NY 1.09   
## 10 7 TX 14.5   
## 11 8 NY 0.832  
## 12 8 TX 14.9   
## 13 9 NY 0.962  
## 14 9 TX 10.7   
## 15 10 NY 1.26   
## 16 10 TX 7.72

# Visualization  
ggplot(TXNY\_rate) +  
 geom\_line(aes(x = Month, y = rate, col = state)) +  
 geom\_point(aes(x = Month, y = rate, col = state))

 ## Compute Total deaths in each state for each month

mydata %>%  
 group\_by(state,Month)%>%  
 summarize(total\_deaths = sum(deathIncrease) )

## `summarise()` regrouping output by 'state' (override with `.groups` argument)

## # A tibble: 457 x 3  
## # Groups: state [56]  
## state Month total\_deaths  
## <chr> <int> <int>  
## 1 AK 3 3  
## 2 AK 4 6  
## 3 AK 5 1  
## 4 AK 6 4  
## 5 AK 7 9  
## 6 AK 8 14  
## 7 AK 9 19  
## 8 AK 10 15  
## 9 AL 3 13  
## 10 AL 4 256  
## # ... with 447 more rows

## Compute the total deaths so far in each state and arrange the rows in the

resulting data frame in (a) descending order of the deaths, (b) ascending order of the deaths

## in descending order of the deaths  
mydata %>%  
 group\_by(state)%>%  
 summarize(death = sum(deathIncrease) )%>%  
 arrange(-death) # or arrange(desc(death))

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

## # A tibble: 56 x 2  
## state death  
## <chr> <int>  
## 1 NY 25773  
## 2 TX 17700  
## 3 CA 17475  
## 4 FL 16775  
## 5 NJ 16324  
## 6 MA 9924  
## 7 IL 9889  
## 8 PA 8718  
## 9 GA 7876  
## 10 MI 7606  
## # ... with 46 more rows

## in ascending order of the deaths  
mydata %>%  
 group\_by(state)%>%  
 summarize(death = sum(deathIncrease) )%>%  
 arrange(death)

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

## # A tibble: 56 x 2  
## state death  
## <chr> <int>  
## 1 AS 0  
## 2 MP 2  
## 3 VI 21  
## 4 VT 58  
## 5 AK 71  
## 6 GU 77  
## 7 WY 77  
## 8 ME 146  
## 9 HI 215  
## 10 MT 325  
## # ... with 46 more rows