Lab 2

Kent Codding

2023-09-07

# Read in Data

setwd("C:/Users/Kent Codding/Desktop/Adv Biostat/Lab 2") #set wd to source file location  
case.data <- read.csv("united\_states\_covid19\_hospitalizations\_deaths\_and\_ed\_visits\_by\_state.csv", skip = 2, row.names = 1)  
vax.data <- read.csv("covid19\_vaccinations\_in\_the\_united\_states.csv" , skip = 5, row.names = 1)

# Operation 1

head(case.data)

## HHS.Region  
## Alabama Region 4  
## Alaska Region 10  
## American Samoa N/A  
## Arizona Region 9  
## Arkansas Region 6  
## California Region 9  
## New.hospital.admissions.of.confirmed.COVID.19.past.week..total.  
## Alabama 374  
## Alaska 21  
## American Samoa 0  
## Arizona 253  
## Arkansas 207  
## California 2593  
## New.COVID.19.hospital.admissions.per.100000.population.past.week..total.  
## Alabama 7.63  
## Alaska 2.87  
## American Samoa 0  
## Arizona 3.48  
## Arkansas 6.86  
## California 6.56  
## X..Change.in.new.hospital.admissions.of.confirmed.COVID.19.from.the.prior.week  
## Alabama 19.5  
## Alaska 5  
## American Samoa 0  
## Arizona 18.2  
## Arkansas 31  
## California 12  
## Total.hospital.admissions.of.confirmed.COVID.19.since.August.1.2020  
## Alabama 116223  
## Alaska 9673  
## American Samoa 0  
## Arizona 153016  
## Arkansas 66900  
## California 603916  
## X..Staffed.inpatient.beds.occupied.by.patients.with.confirmed.COVID.19.past.week..average.  
## Alabama 3.1  
## Alaska 1.3  
## American Samoa 0  
## Arizona 1.5  
## Arkansas 2.4  
## California 3.8  
## Absolute.change...staffed.inpatient.beds.occupied.by.patients.with.confirmed.COVID.19.from.prior.week  
## Alabama 0.5  
## Alaska 0.1  
## American Samoa 0  
## Arizona 0.3  
## Arkansas 0.6  
## California 0.6  
## X..Staffed.ICU.beds.occupied.by.patients.with.confirmed.COVID.19.past.week..average.  
## Alabama 2.7  
## Alaska 0.6  
## American Samoa 0  
## Arizona 0.8  
## Arkansas 1.6  
## California 1.9  
## Absolute.change...staffed.ICU.beds.occupied.by.patients.with.confirmed.COVID.19.from.prior.week  
## Alabama 0.5  
## Alaska 0.3  
## American Samoa 0  
## Arizona 0.2  
## Arkansas 0.7  
## California 0.2  
## Percentage.of.deaths.due.to.COVID.19.in.past.week  
## Alabama 0  
## Alaska 0  
## American Samoa Data not available  
## Arizona Counts 1-9  
## Arkansas Counts 1-9  
## California 2.3  
## Percent.change.percentage.of.deaths.due.to.COVID.19.in.past.week  
## Alabama -100  
## Alaska N/A  
## American Samoa N/A  
## Arizona N/A  
## Arkansas N/A  
## California 9.5  
## Absolute.change.percentage.of.deaths.due.to.COVID.19.in.past.week  
## Alabama -2.5  
## Alaska 0  
## American Samoa N/A  
## Arizona N/A  
## Arkansas N/A  
## California 0.2  
## Deaths.in.the.Past.3.Months  
## Alabama 127  
## Alaska 17  
## American Samoa Data not available  
## Arizona 167  
## Arkansas 68  
## California 718  
## Death.Rate.per.100000.for.the.Past.3.Months Total.Deaths  
## Alabama 2 22342  
## Alaska Counts 1-9 1514  
## American Samoa Data not available Data not available  
## Arizona 1.8 29685  
## Arkansas 1.9 12531  
## California 1.7 108732  
## Total.Death.rate.per.100000  
## Alabama 358.8  
## Alaska 226.9  
## American Samoa Data not available  
## Arizona 324.5  
## Arkansas 337.5  
## California 245.3  
## Percent.of.ED.visits.diagnosed.as.COVID.19  
## Alabama 4.5%  
## Alaska 1.8%  
## American Samoa N/A  
## Arizona 1.5%  
## Arkansas 2.3%  
## California 2.1%  
## Percent.change.in...of.visits.diagnosed.as.COVID.19.from.last.week.compared.to.prior.week  
## Alabama 22.3%  
## Alaska 16.1%  
## American Samoa N/A  
## Arizona -14.3%  
## Arkansas 29.6%  
## California 1.1%  
## Test.positivity.....in.past.week  
## Alabama 13.1  
## Alaska 12.9  
## American Samoa N/A  
## Arizona 16.1  
## Arkansas 20.8  
## California 16.1  
## X..Change.in.test.positivity.compared.with.prior.week  
## Alabama -0.7  
## Alaska 0.7  
## American Samoa N/A  
## Arizona 2.6  
## Arkansas 2.4  
## California 2.6  
## Test.positivity.....in.past.2.weeks  
## Alabama 13.2  
## Alaska 12.3  
## American Samoa N/A  
## Arizona 13.8  
## Arkansas 17.8  
## California 13.8  
## Test.positivity.....in.past.4.weeks Test.volume.in.past.week  
## Alabama 11.8 4012  
## Alaska 11.1 5912  
## American Samoa N/A N/A  
## Arizona 12.3 2799  
## Arkansas 16.3 9749  
## California 12.3 2799  
## Test.volume.in.past.2.weeks Test.volume.in.past.4.weeks  
## Alabama 9502 19072  
## Alaska 11277 21556  
## American Samoa N/A N/A  
## Arizona 9368 19354  
## Arkansas 18409 30581  
## California 9368 19354

# keep certain names

names(case.data)

## [1] "HHS.Region"   
## [2] "New.hospital.admissions.of.confirmed.COVID.19.past.week..total."   
## [3] "New.COVID.19.hospital.admissions.per.100000.population.past.week..total."   
## [4] "X..Change.in.new.hospital.admissions.of.confirmed.COVID.19.from.the.prior.week"   
## [5] "Total.hospital.admissions.of.confirmed.COVID.19.since.August.1.2020"   
## [6] "X..Staffed.inpatient.beds.occupied.by.patients.with.confirmed.COVID.19.past.week..average."   
## [7] "Absolute.change...staffed.inpatient.beds.occupied.by.patients.with.confirmed.COVID.19.from.prior.week"  
## [8] "X..Staffed.ICU.beds.occupied.by.patients.with.confirmed.COVID.19.past.week..average."   
## [9] "Absolute.change...staffed.ICU.beds.occupied.by.patients.with.confirmed.COVID.19.from.prior.week"   
## [10] "Percentage.of.deaths.due.to.COVID.19.in.past.week"   
## [11] "Percent.change.percentage.of.deaths.due.to.COVID.19.in.past.week"   
## [12] "Absolute.change.percentage.of.deaths.due.to.COVID.19.in.past.week"   
## [13] "Deaths.in.the.Past.3.Months"   
## [14] "Death.Rate.per.100000.for.the.Past.3.Months"   
## [15] "Total.Deaths"   
## [16] "Total.Death.rate.per.100000"   
## [17] "Percent.of.ED.visits.diagnosed.as.COVID.19"   
## [18] "Percent.change.in...of.visits.diagnosed.as.COVID.19.from.last.week.compared.to.prior.week"   
## [19] "Test.positivity.....in.past.week"   
## [20] "X..Change.in.test.positivity.compared.with.prior.week"   
## [21] "Test.positivity.....in.past.2.weeks"   
## [22] "Test.positivity.....in.past.4.weeks"   
## [23] "Test.volume.in.past.week"   
## [24] "Test.volume.in.past.2.weeks"   
## [25] "Test.volume.in.past.4.weeks"

case.data <- case.data[,c("Total.Deaths", "Death.Rate.per.100000.for.the.Past.3.Months", "Total.Death.rate.per.100000", "New.COVID.19.hospital.admissions.per.100000.population.past.week..total.")]   
head(case.data)

## Total.Deaths Death.Rate.per.100000.for.the.Past.3.Months  
## Alabama 22342 2  
## Alaska 1514 Counts 1-9  
## American Samoa Data not available Data not available  
## Arizona 29685 1.8  
## Arkansas 12531 1.9  
## California 108732 1.7  
## Total.Death.rate.per.100000  
## Alabama 358.8  
## Alaska 226.9  
## American Samoa Data not available  
## Arizona 324.5  
## Arkansas 337.5  
## California 245.3  
## New.COVID.19.hospital.admissions.per.100000.population.past.week..total.  
## Alabama 7.63  
## Alaska 2.87  
## American Samoa 0  
## Arizona 3.48  
## Arkansas 6.86  
## California 6.56

# Operation 2

names(vax.data)

## [1] "Total.doses.distributed"   
## [2] "Doses.distributed.per.100k.pop"   
## [3] "Doses.distributed.by.jurisdiction.per.100k.of.18..pop"   
## [4] "Total.doses.administered.by.jurisdiction"   
## [5] "Doses.administered.by.jurisdiction.per.100k.pop"   
## [6] "Doses.administered.by.jurisdiction.to.18..pop"   
## [7] "Doses.administered.by.jurisdiction.per.100k.of.18..pop"   
## [8] "Residents.with.at.least.one.dose"   
## [9] "Percent.of.total.pop.with.at.least.one.dose"   
## [10] "Residents.18..with.at.least.one.dose"   
## [11] "Percent.of.18..pop.with.at.least.one.dose"   
## [12] "Residents.with.a.completed.primary.series"   
## [13] "Percent.of.total.pop.with.a.completed.primary.series"   
## [14] "Residents.18..with.a.completed.primary.series"   
## [15] "Percent.of.18..pop.with.a.completed.primary.series"   
## [16] "Total.number.of.original.Pfizer.doses.distributed"   
## [17] "Total.number.of.Pfizer.updated.booster.doses.distributed"   
## [18] "Total.number.of.original.Moderna.doses.distributed"   
## [19] "Total.number.of.Moderna.updated.booster.doses.distributed"   
## [20] "Total.number.of.Janssen.doses.distributed"   
## [21] "Total.number.of.Novavax.doses.distributed"   
## [22] "Total.number.of.doses.from.other.manufacturer.distributed"   
## [23] "Total.number.of.Janssen.doses.administered"   
## [24] "Total.number.of.original.Moderna.doses.administered"   
## [25] "Total.number.of.Moderna.updated.booster.doses.administered"   
## [26] "Total.number.of.original.Pfizer.doses.administered"   
## [27] "Total.number.of.Pfizer.updated.booster.doses.administered"   
## [28] "Total.number.of.Novavax.doses.administered"   
## [29] "Total.number.of.doses.from.other.manufacturer.administered"   
## [30] "Residents.65..with.at.least.one.dose"   
## [31] "Percent.of.65..pop.with.at.least.one.dose"   
## [32] "Residents.65..with.a.completed.primary.series"   
## [33] "Percent.of.65..pop.with.a.completed.primary.series"   
## [34] "Doses.administered.by.jurisdiction.to.65..pop"   
## [35] "Doses.administered.by.jurisdiction.per.100k.of.65..pop"   
## [36] "Doses.distributed.by.jurisdiction.per.100k.of.65..pop"   
## [37] "Residents.12..with.at.least.one.dose"   
## [38] "Percent.of.12..pop.with.at.least.one.dose"   
## [39] "Residents.12..with.a.completed.primary.series"   
## [40] "Percent.of.12..pop.with.a.completed.primary.series"   
## [41] "Doses.administered.by.jurisdiction.to.12..pop"   
## [42] "Doses.administered.by.jurisdiction.per.100k.of.12..pop"   
## [43] "Doses.distributed.by.jurisdiction.per.100k.of.12..pop"   
## [44] "Residents.5..with.at.least.one.dose"   
## [45] "Percent.of.5..pop.with.at.least.one.dose"   
## [46] "Residents.5..with.a.completed.primary.series"   
## [47] "Percent.of.5..pop.with.a.completed.primary.series"   
## [48] "Doses.administered.by.jurisdiction.to.5..pop"   
## [49] "Doses.administered.by.jurisdiction.per.100k.of.5..pop"   
## [50] "Doses.distributed.by.jurisdiction.per.100k.of.5..pop"   
## [51] "Residents.with.an.updated..bivalent..booster.dose"   
## [52] "Percent.of.pop.with.an.updated..bivalent..booster.dose"   
## [53] "Residents.5..with.an.updated..bivalent..booster.dose"   
## [54] "Percent.of.5..pop.with.an.updated..bivalent..booster.dose"   
## [55] "Residents.12..with.an.updated..bivalent..booster.dose"   
## [56] "Percent.of.12..pop.with.an.updated..bivalent..booster.dose"   
## [57] "Residents.18..with.an.updated..bivalent..booster.dose"   
## [58] "Percent.of.18..pop.with.an.updated..bivalent..booster.dose"   
## [59] "Residents.65..with.an.updated..bivalent..booster.dose"   
## [60] "Percent.of.65..pop.with.an.updated..bivalent..booster.dose"   
## [61] "Children..5.with.at.least.one.dose"   
## [62] "Total.number.of.updated..bivalent..booster.doses.administered"

vax.data <- vax.data[,c("Percent.of.total.pop.with.at.least.one.dose", "Percent.of.total.pop.with.a.completed.primary.series")]   
head(vax.data)

## Percent.of.total.pop.with.at.least.one.dose  
## United States 81.4  
## Alaska 73.2  
## Alabama 65.1  
## Arkansas 70.1  
## American Samoa 95  
## Arizona 78.4  
## Percent.of.total.pop.with.a.completed.primary.series  
## United States 69.5  
## Alaska 65.3  
## Alabama 53.3  
## Arkansas 57  
## American Samoa 89.7  
## Arizona 66.2

# Merge the data

state.name[!(state.name %in% rownames(case.data))]

## character(0)

# Operation 3

state.name[!(state.name %in% rownames(vax.data))]

## [1] "New York"

rownames(vax.data)

## [1] "United States" "Alaska"   
## [3] "Alabama" "Arkansas"   
## [5] "American Samoa" "Arizona"   
## [7] "Bureau of Prisons" "California"   
## [9] "Colorado" "Connecticut"   
## [11] "District of Columbia" "Dept of Defense"   
## [13] "Delaware" "Florida"   
## [15] "Federated States of Micronesia" "Georgia"   
## [17] "Guam" "Hawaii"   
## [19] "Iowa" "Idaho"   
## [21] "Indian Health Svc" "Illinois"   
## [23] "Indiana" "Kansas"   
## [25] "Kentucky" "Louisiana"   
## [27] "Massachusetts" "Maryland"   
## [29] "Maine" "Marshall Islands"   
## [31] "Michigan" "Minnesota"   
## [33] "Missouri" "Northern Mariana Islands"   
## [35] "Mississippi" "Montana"   
## [37] "North Carolina" "North Dakota"   
## [39] "Nebraska" "New Hampshire"   
## [41] "New Jersey" "New Mexico"   
## [43] "Nevada" "New York State"   
## [45] "Ohio" "Oklahoma"   
## [47] "Oregon" "Pennsylvania"   
## [49] "Puerto Rico" "Republic of Palau"   
## [51] "Rhode Island" "South Carolina"   
## [53] "South Dakota" "Tennessee"   
## [55] "Texas" "Utah"   
## [57] "Virginia" "Veterans Health"   
## [59] "Virgin Islands" "Vermont"   
## [61] "Washington" "Wisconsin"   
## [63] "West Virginia" "Wyoming"

# Question 1

at index 44 in the output from rownames(vax.data) above, New York is listed as “New York State.” Thus, this line of code - state.name[!(state.name %in% rownames(vax.data))] - that checks for state names where the state name Character Object is NOT in the rownames of vax data will return “New York” because R reads “New York State” as a different character object than “New York.”

# Question 4

rownames(vax.data)[44] <- "New York"  
row.names(vax.data)

## [1] "United States" "Alaska"   
## [3] "Alabama" "Arkansas"   
## [5] "American Samoa" "Arizona"   
## [7] "Bureau of Prisons" "California"   
## [9] "Colorado" "Connecticut"   
## [11] "District of Columbia" "Dept of Defense"   
## [13] "Delaware" "Florida"   
## [15] "Federated States of Micronesia" "Georgia"   
## [17] "Guam" "Hawaii"   
## [19] "Iowa" "Idaho"   
## [21] "Indian Health Svc" "Illinois"   
## [23] "Indiana" "Kansas"   
## [25] "Kentucky" "Louisiana"   
## [27] "Massachusetts" "Maryland"   
## [29] "Maine" "Marshall Islands"   
## [31] "Michigan" "Minnesota"   
## [33] "Missouri" "Northern Mariana Islands"   
## [35] "Mississippi" "Montana"   
## [37] "North Carolina" "North Dakota"   
## [39] "Nebraska" "New Hampshire"   
## [41] "New Jersey" "New Mexico"   
## [43] "Nevada" "New York"   
## [45] "Ohio" "Oklahoma"   
## [47] "Oregon" "Pennsylvania"   
## [49] "Puerto Rico" "Republic of Palau"   
## [51] "Rhode Island" "South Carolina"   
## [53] "South Dakota" "Tennessee"   
## [55] "Texas" "Utah"   
## [57] "Virginia" "Veterans Health"   
## [59] "Virgin Islands" "Vermont"   
## [61] "Washington" "Wisconsin"   
## [63] "West Virginia" "Wyoming"

# Operation 5

all.data <- merge(vax.data, case.data, by = 0) #merges vax.data and case.data by row

# Operation 6

all.data <- data.frame(all.data, row.names = all.data$Row.names) #removes numerical row names (indices) and replaces with state names   
all.data <- all.data[,-1] #gets rid of first column that contained redundant row names  
rownames(all.data)

## [1] "Alabama" "Alaska"   
## [3] "American Samoa" "Arizona"   
## [5] "Arkansas" "California"   
## [7] "Colorado" "Connecticut"   
## [9] "Delaware" "District of Columbia"   
## [11] "Federated States of Micronesia" "Florida"   
## [13] "Georgia" "Guam"   
## [15] "Hawaii" "Idaho"   
## [17] "Illinois" "Indiana"   
## [19] "Iowa" "Kansas"   
## [21] "Kentucky" "Louisiana"   
## [23] "Maine" "Maryland"   
## [25] "Massachusetts" "Michigan"   
## [27] "Minnesota" "Mississippi"   
## [29] "Missouri" "Montana"   
## [31] "Nebraska" "Nevada"   
## [33] "New Hampshire" "New Jersey"   
## [35] "New Mexico" "New York"   
## [37] "North Carolina" "North Dakota"   
## [39] "Northern Mariana Islands" "Ohio"   
## [41] "Oklahoma" "Oregon"   
## [43] "Pennsylvania" "Puerto Rico"   
## [45] "Rhode Island" "South Carolina"   
## [47] "South Dakota" "Tennessee"   
## [49] "Texas" "Utah"   
## [51] "Vermont" "Virgin Islands"   
## [53] "Virginia" "Washington"   
## [55] "West Virginia" "Wisconsin"   
## [57] "Wyoming"

# Operation 7

library(magrittr) # load magrittr for pipe operator  
subset(all.data, row.names(all.data) %in% state.name) %>% row.names() #check if this subset command only includes 50 states as rownames

## [1] "Alabama" "Alaska" "Arizona" "Arkansas"   
## [5] "California" "Colorado" "Connecticut" "Delaware"   
## [9] "Florida" "Georgia" "Hawaii" "Idaho"   
## [13] "Illinois" "Indiana" "Iowa" "Kansas"   
## [17] "Kentucky" "Louisiana" "Maine" "Maryland"   
## [21] "Massachusetts" "Michigan" "Minnesota" "Mississippi"   
## [25] "Missouri" "Montana" "Nebraska" "Nevada"   
## [29] "New Hampshire" "New Jersey" "New Mexico" "New York"   
## [33] "North Carolina" "North Dakota" "Ohio" "Oklahoma"   
## [37] "Oregon" "Pennsylvania" "Rhode Island" "South Carolina"  
## [41] "South Dakota" "Tennessee" "Texas" "Utah"   
## [45] "Vermont" "Virginia" "Washington" "West Virginia"   
## [49] "Wisconsin" "Wyoming"

all.data <- subset(all.data, row.names(all.data) %in% state.name) #get rid of territories  
names(all.data) <- c("perc\_one\_dose","perc\_fully\_vax",  
 "total\_deaths","death\_rate\_3\_months",  
 "total\_death\_rate","new\_case\_rate\_this\_week") #make better column names

# Operation 8

str(all.data)

## 'data.frame': 50 obs. of 6 variables:  
## $ perc\_one\_dose : chr "65.1" "73.2" "78.4" "70.1" ...  
## $ perc\_fully\_vax : chr "53.3" "65.3" "66.2" "57" ...  
## $ total\_deaths : chr "22342" "1514" "29685" "12531" ...  
## $ death\_rate\_3\_months : chr "2" "Counts 1-9" "1.8" "1.9" ...  
## $ total\_death\_rate : chr "358.8" "226.9" "324.5" "337.5" ...  
## $ new\_case\_rate\_this\_week: chr "7.63" "2.87" "3.48" "6.86" ...

all.data <- data.frame(lapply(all.data, as.numeric),  
 row.names = rownames(all.data))

## Warning in lapply(all.data, as.numeric): NAs introduced by coercion

# Question 2

R gave us the warning message “NAs introduced by coercion” because some character values like “Data not available” found in all.data cannot be converted to a number. Instead, R coerces these values to NAs.

# Read in usdata

library(usdata)

## Warning: package 'usdata' was built under R version 4.1.3

data(state\_stats)  
state\_stats <- data.frame(state\_stats, row.names = 1)

# Operation 9

all.data <- merge(all.data, state\_stats, by = 0) #merges vax.data and case.data by row  
head(all.data)

## Row.names perc\_one\_dose perc\_fully\_vax total\_deaths death\_rate\_3\_months  
## 1 Alabama 65.1 53.3 22342 2.0  
## 2 Alaska 73.2 65.3 1514 NA  
## 3 Arizona 78.4 66.2 29685 1.8  
## 4 Arkansas 70.1 57.0 12531 1.9  
## 5 California 85.1 74.9 108732 1.7  
## 6 Colorado 84.0 73.8 14953 1.6  
## total\_death\_rate new\_case\_rate\_this\_week abbr fips pop2010 pop2000  
## 1 358.8 7.63 AL 1 4779736 4447100  
## 2 226.9 2.87 AK 2 710231 626932  
## 3 324.5 3.48 AZ 4 6392017 5130632  
## 4 337.5 6.86 AR 5 2915918 2673400  
## 5 245.3 6.56 CA 6 37253956 33871648  
## 6 246.6 3.32 CO 8 5029196 4301261  
## homeownership multiunit income med\_income poverty fed\_spend land\_area smoke  
## 1 71.1 15.5 22984 42081 17.1 11.726134 50645.33 24.8  
## 2 64.7 24.6 30726 66521 9.5 16.786568 570640.95 25.0  
## 3 67.4 20.7 25680 50448 15.3 9.847128 113594.08 20.4  
## 4 67.7 15.2 21274 39267 18.0 9.611186 52035.48 23.5  
## 5 57.4 30.7 29188 60883 13.7 8.885791 155779.22 15.2  
## 6 67.6 25.6 30151 56456 12.2 9.152295 103641.89 19.9  
## murder robbery agg\_assault larceny motor\_theft soc\_sec nuclear coal tr\_deaths  
## 1 8.2 141.4 5.2 55.4 6.0 18.5 16.3 56.6 23.7  
## 2 4.8 80.9 65.5 366.0 55.1 8.9 0.0 9.5 10.1  
## 3 7.5 144.4 5.1 46.4 14.5 13.9 15.6 39.6 18.4  
## 4 6.7 91.1 13.3 93.0 9.0 18.7 13.0 48.2 22.2  
## 5 6.9 176.1 0.9 5.1 1.9 11.8 7.0 1.1 11.6  
## 6 3.7 84.6 5.3 54.4 11.1 11.4 0.0 71.7 12.0  
## tr\_deaths\_no\_alc unempl  
## 1 1.3 7.6  
## 2 7.3 7.1  
## 3 0.9 8.7  
## 4 2.2 7.6  
## 5 0.2 10.9  
## 6 1.2 7.8

## write to csv

write.csv(all.data, file = "all\_data.csv")