Homework\_3

Sijay Matsinye

2/16/2021

##1) Enter and run USArrests. What type of information is shown in the data table USArrests? (Three or four sentences)

\*\* This data table shows the rates of arrests for three different types of violent crimes broken down by state. THe table includes four variable that are the three crimes: murder,assault, rape along with the percent of the population living in urban areas. Lastly the rate represented for each crime is calculated per every 100,000 residents.

library(tidyverse)

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

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.3 v dplyr 1.0.2  
## 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()

USArrests

## Murder Assault UrbanPop Rape  
## Alabama 13.2 236 58 21.2  
## Alaska 10.0 263 48 44.5  
## Arizona 8.1 294 80 31.0  
## Arkansas 8.8 190 50 19.5  
## California 9.0 276 91 40.6  
## Colorado 7.9 204 78 38.7  
## Connecticut 3.3 110 77 11.1  
## Delaware 5.9 238 72 15.8  
## Florida 15.4 335 80 31.9  
## Georgia 17.4 211 60 25.8  
## Hawaii 5.3 46 83 20.2  
## Idaho 2.6 120 54 14.2  
## Illinois 10.4 249 83 24.0  
## Indiana 7.2 113 65 21.0  
## Iowa 2.2 56 57 11.3  
## Kansas 6.0 115 66 18.0  
## Kentucky 9.7 109 52 16.3  
## Louisiana 15.4 249 66 22.2  
## Maine 2.1 83 51 7.8  
## Maryland 11.3 300 67 27.8  
## Massachusetts 4.4 149 85 16.3  
## Michigan 12.1 255 74 35.1  
## Minnesota 2.7 72 66 14.9  
## Mississippi 16.1 259 44 17.1  
## Missouri 9.0 178 70 28.2  
## Montana 6.0 109 53 16.4  
## Nebraska 4.3 102 62 16.5  
## Nevada 12.2 252 81 46.0  
## New Hampshire 2.1 57 56 9.5  
## New Jersey 7.4 159 89 18.8  
## New Mexico 11.4 285 70 32.1  
## New York 11.1 254 86 26.1  
## North Carolina 13.0 337 45 16.1  
## North Dakota 0.8 45 44 7.3  
## Ohio 7.3 120 75 21.4  
## Oklahoma 6.6 151 68 20.0  
## Oregon 4.9 159 67 29.3  
## Pennsylvania 6.3 106 72 14.9  
## Rhode Island 3.4 174 87 8.3  
## South Carolina 14.4 279 48 22.5  
## South Dakota 3.8 86 45 12.8  
## Tennessee 13.2 188 59 26.9  
## Texas 12.7 201 80 25.5  
## Utah 3.2 120 80 22.9  
## Vermont 2.2 48 32 11.2  
## Virginia 8.5 156 63 20.7  
## Washington 4.0 145 73 26.2  
## West Virginia 5.7 81 39 9.3  
## Wisconsin 2.6 53 66 10.8  
## Wyoming 6.8 161 60 15.6

##2) Use and show R coding that features a map function to show maximum values for all variables of the USArerests data frame. Which State has the largest number of Assaults according to the USAressts data frame ?

\*\*North Carolina has the largest number of Assaults according to the USArrests data frame.

map\_dbl(USArrests, max)

## Murder Assault UrbanPop Rape   
## 17.4 337.0 91.0 46.0

USArrests %>%   
 arrange(desc(Assault))

## Murder Assault UrbanPop Rape  
## North Carolina 13.0 337 45 16.1  
## Florida 15.4 335 80 31.9  
## Maryland 11.3 300 67 27.8  
## Arizona 8.1 294 80 31.0  
## New Mexico 11.4 285 70 32.1  
## South Carolina 14.4 279 48 22.5  
## California 9.0 276 91 40.6  
## Alaska 10.0 263 48 44.5  
## Mississippi 16.1 259 44 17.1  
## Michigan 12.1 255 74 35.1  
## New York 11.1 254 86 26.1  
## Nevada 12.2 252 81 46.0  
## Illinois 10.4 249 83 24.0  
## Louisiana 15.4 249 66 22.2  
## Delaware 5.9 238 72 15.8  
## Alabama 13.2 236 58 21.2  
## Georgia 17.4 211 60 25.8  
## Colorado 7.9 204 78 38.7  
## Texas 12.7 201 80 25.5  
## Arkansas 8.8 190 50 19.5  
## Tennessee 13.2 188 59 26.9  
## Missouri 9.0 178 70 28.2  
## Rhode Island 3.4 174 87 8.3  
## Wyoming 6.8 161 60 15.6  
## New Jersey 7.4 159 89 18.8  
## Oregon 4.9 159 67 29.3  
## Virginia 8.5 156 63 20.7  
## Oklahoma 6.6 151 68 20.0  
## Massachusetts 4.4 149 85 16.3  
## Washington 4.0 145 73 26.2  
## Idaho 2.6 120 54 14.2  
## Ohio 7.3 120 75 21.4  
## Utah 3.2 120 80 22.9  
## Kansas 6.0 115 66 18.0  
## Indiana 7.2 113 65 21.0  
## Connecticut 3.3 110 77 11.1  
## Kentucky 9.7 109 52 16.3  
## Montana 6.0 109 53 16.4  
## Pennsylvania 6.3 106 72 14.9  
## Nebraska 4.3 102 62 16.5  
## South Dakota 3.8 86 45 12.8  
## Maine 2.1 83 51 7.8  
## West Virginia 5.7 81 39 9.3  
## Minnesota 2.7 72 66 14.9  
## New Hampshire 2.1 57 56 9.5  
## Iowa 2.2 56 57 11.3  
## Wisconsin 2.6 53 66 10.8  
## Vermont 2.2 48 32 11.2  
## Hawaii 5.3 46 83 20.2  
## North Dakota 0.8 45 44 7.3

##3) Install the nycflights13 package: install.packages(“nycflights13”), call the following library:library(nycflights13), and then enter (this will produce the flights data table)

#install.packages("nycflights13")  
library(nycflights13)

## Warning: package 'nycflights13' was built under R version 4.0.3

flights

## # A tibble: 336,776 x 19  
## year month day dep\_time sched\_dep\_time dep\_delay arr\_time sched\_arr\_time  
## <int> <int> <int> <int> <int> <dbl> <int> <int>  
## 1 2013 1 1 517 515 2 830 819  
## 2 2013 1 1 533 529 4 850 830  
## 3 2013 1 1 542 540 2 923 850  
## 4 2013 1 1 544 545 -1 1004 1022  
## 5 2013 1 1 554 600 -6 812 837  
## 6 2013 1 1 554 558 -4 740 728  
## 7 2013 1 1 555 600 -5 913 854  
## 8 2013 1 1 557 600 -3 709 723  
## 9 2013 1 1 557 600 -3 838 846  
## 10 2013 1 1 558 600 -2 753 745  
## # ... with 336,766 more rows, and 11 more variables: arr\_delay <dbl>,  
## # carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,  
## # air\_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time\_hour <dttm>

##4) Use and show R code that will indicate how many rows and how many columns the flights data has. Review and revisit your notes from STAT 412/612. Describe a tibble (two or three sentences).Now use and show R code that verifies that flights is a tibble

\*\*A tibble is a data frame that doesn't perform al of the actions of a data frame. When printed it only shows a preview of all the values and objects in the data frame and it can't change the object types. Overall it performs modified actions of that which a data frame does.

flights %>%   
 nrow()

## [1] 336776

flights %>%   
 ncol()

## [1] 19

flights %>%   
 is\_tibble()

## [1] TRUE

##5)Now Use and show R code (featuring a map function) that will output the type of each column of the flights tibble.

flights

## # A tibble: 336,776 x 19  
## year month day dep\_time sched\_dep\_time dep\_delay arr\_time sched\_arr\_time  
## <int> <int> <int> <int> <int> <dbl> <int> <int>  
## 1 2013 1 1 517 515 2 830 819  
## 2 2013 1 1 533 529 4 850 830  
## 3 2013 1 1 542 540 2 923 850  
## 4 2013 1 1 544 545 -1 1004 1022  
## 5 2013 1 1 554 600 -6 812 837  
## 6 2013 1 1 554 558 -4 740 728  
## 7 2013 1 1 555 600 -5 913 854  
## 8 2013 1 1 557 600 -3 709 723  
## 9 2013 1 1 557 600 -3 838 846  
## 10 2013 1 1 558 600 -2 753 745  
## # ... with 336,766 more rows, and 11 more variables: arr\_delay <dbl>,  
## # carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,  
## # air\_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time\_hour <dttm>

map(flights,typeof)

## $year  
## [1] "integer"  
##   
## $month  
## [1] "integer"  
##   
## $day  
## [1] "integer"  
##   
## $dep\_time  
## [1] "integer"  
##   
## $sched\_dep\_time  
## [1] "integer"  
##   
## $dep\_delay  
## [1] "double"  
##   
## $arr\_time  
## [1] "integer"  
##   
## $sched\_arr\_time  
## [1] "integer"  
##   
## $arr\_delay  
## [1] "double"  
##   
## $carrier  
## [1] "character"  
##   
## $flight  
## [1] "integer"  
##   
## $tailnum  
## [1] "character"  
##   
## $origin  
## [1] "character"  
##   
## $dest  
## [1] "character"  
##   
## $air\_time  
## [1] "double"  
##   
## $distance  
## [1] "double"  
##   
## $hour  
## [1] "double"  
##   
## $minute  
## [1] "double"  
##   
## $time\_hour  
## [1] "double"

##6)Use and show R coding that features usage of a map function to find the slope and the intercept of models for the different levels of the Species variable of the iris data frame. For each model, Sepal.Width predicts Sepal.Length.

models <- iris %>%  
 split(.$Species) %>%  
 map(~lm(Sepal.Length ~ Sepal.Width, data = .))  
models

## $setosa  
##   
## Call:  
## lm(formula = Sepal.Length ~ Sepal.Width, data = .)  
##   
## Coefficients:  
## (Intercept) Sepal.Width   
## 2.6390 0.6905   
##   
##   
## $versicolor  
##   
## Call:  
## lm(formula = Sepal.Length ~ Sepal.Width, data = .)  
##   
## Coefficients:  
## (Intercept) Sepal.Width   
## 3.5397 0.8651   
##   
##   
## $virginica  
##   
## Call:  
## lm(formula = Sepal.Length ~ Sepal.Width, data = .)  
##   
## Coefficients:  
## (Intercept) Sepal.Width   
## 3.9068 0.9015

V <- list(12, 22, 27, 31.5, NA, 39, "east")  
V

## [[1]]  
## [1] 12  
##   
## [[2]]  
## [1] 22  
##   
## [[3]]  
## [1] 27  
##   
## [[4]]  
## [1] 31.5  
##   
## [[5]]  
## [1] NA  
##   
## [[6]]  
## [1] 39  
##   
## [[7]]  
## [1] "east"

7a) Use and show R code to find the length of the list

length(V)

## [1] 7

7b) Use and show R code that will extract the missing value

V[is.na(V)]

## [[1]]  
## [1] NA

7c) Use and show R code that will extract the third object

V[3]

## [[1]]  
## [1] 27

7d) Use and show R code that will extract the character string and the minimum number.

str(V[1])

## List of 1  
## $ : num 12

str(V[7])

## List of 1  
## $ : chr "east"

7e) Use one line of code to show that the seventh object is a character object.

str(V[[7]])

## chr "east"