HW 2

Lindsay Beyak

2/19/2021

#### Exercise 1

The USArrests dataset contains arrests that occurred in the USA in 1973 for assault, murder, and rape.Each of the 50 states are listed with 4 columns measured numerically. The ‘Murder’, ‘Arrest’, and Rape columns indicate arrests (per 100,000), while the ‘UrbanPop’ column represents the percentage urban population.

library(tidyverse)

## ── Attaching packages ────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.2 ✓ purrr 0.3.4  
## ✓ tibble 3.0.3 ✓ dplyr 1.0.2  
## ✓ tidyr 1.1.1 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.5.0

## ── Conflicts ───────────────────────────────────────────── tidyverse\_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

typeof(USArrests)

## [1] "list"

length(USArrests)

## [1] 4

dim(USArrests)

## [1] 50 4

#### Exercise 2

Based on the following code, the state with the most Assaults is North Carolina with 337.

map\_dbl(USArrests, max)

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

#### Exercise 3

library(nycflights13)  
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>

#### Exercise 4

A tibble is a data frame that is more unified than a traditional R data frame. Tibble is part of the tidyverse package. It is a more organized data frame to give a cleaner overview of the columns and rows.

#counts for length of columns and rows  
length(flights)

## [1] 19

nrow(flights)

## [1] 336776

ncol(flights)

## [1] 19

#Tibble confirmations for flights  
count(flights)

## # A tibble: 1 x 1  
## n  
## <int>  
## 1 336776

is\_tibble(flights)

## [1] TRUE

class(flights)

## [1] "tbl\_df" "tbl" "data.frame"

#### Exercise 5

Below shows the code to reveal each column type of the flights tibble.

typeflcol <- vector("list", ncol(nycflights13::flights))  
names(typeflcol) <- names(nycflights13::flights)  
for (i in names(nycflights13::flights)) {  
 typeflcol[[i]] <- class(nycflights13::flights[[i]])  
}  
typeflcol

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

#### Exercise 6

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

## $setosa  
##   
## Call:  
## lm(formula = Sepal.Width ~ Sepal.Length, data = .)  
##   
## Coefficients:  
## (Intercept) Sepal.Length   
## -0.5694 0.7985   
##   
##   
## $versicolor  
##   
## Call:  
## lm(formula = Sepal.Width ~ Sepal.Length, data = .)  
##   
## Coefficients:  
## (Intercept) Sepal.Length   
## 0.8721 0.3197   
##   
##   
## $virginica  
##   
## Call:  
## lm(formula = Sepal.Width ~ Sepal.Length, data = .)  
##   
## Coefficients:  
## (Intercept) Sepal.Length   
## 1.4463 0.2319

#### Exercise 7

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 Length of the list  
length(V)

## [1] 7

#7b Extract the missing value  
V[is.na(V)]

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

#7c Third object extracted  
V[3]

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

#7d Extracting the character string and the minimum number  
V[min(1:6)]

## [[1]]  
## [1] 12

V[7]

## [[1]]  
## [1] "east"

#7e 7th line of code is a character object  
str(V)

## List of 7  
## $ : num 12  
## $ : num 22  
## $ : num 27  
## $ : num 31.5  
## $ : logi NA  
## $ : num 39  
## $ : chr "east"