DS hw 2

Sarah Hassan

2/19/2021

library(tidyverse)

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

## ✓ ggplot2 3.3.2 ✓ purrr 0.3.4  
## ✓ tibble 3.0.4 ✓ dplyr 1.0.2  
## ✓ tidyr 1.1.2 ✓ 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()

library(dplyr)  
library(nycflights13)  
  
#1  
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

#In this data set we see crime rates for each state, containing 4 different variables. In this case the variables are murder, assult, rape, and urban population percentages.   
  
#2  
USArrests %>%  
 map\_dbl(max)

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

#North Carolina has the largest number of assaults  
  
#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  
nrow(flights)

## [1] 336776

ncol(flights)

## [1] 19

#tibbles are a type of data set that displays just the first 10 rows, unlike a regular dataset. The input type also always stays the same.   
  
is\_tibble(flights)

## [1] TRUE

#5  
  
flights %>%  
 map(summary)

## $year  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2013 2013 2013 2013 2013 2013   
##   
## $month  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 4.000 7.000 6.549 10.000 12.000   
##   
## $day  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 8.00 16.00 15.71 23.00 31.00   
##   
## $dep\_time  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1 907 1401 1349 1744 2400 8255   
##   
## $sched\_dep\_time  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 106 906 1359 1344 1729 2359   
##   
## $dep\_delay  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -43.00 -5.00 -2.00 12.64 11.00 1301.00 8255   
##   
## $arr\_time  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1 1104 1535 1502 1940 2400 8713   
##   
## $sched\_arr\_time  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1 1124 1556 1536 1945 2359   
##   
## $arr\_delay  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -86.000 -17.000 -5.000 6.895 14.000 1272.000 9430   
##   
## $carrier  
## Length Class Mode   
## 336776 character character   
##   
## $flight  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1 553 1496 1972 3465 8500   
##   
## $tailnum  
## Length Class Mode   
## 336776 character character   
##   
## $origin  
## Length Class Mode   
## 336776 character character   
##   
## $dest  
## Length Class Mode   
## 336776 character character   
##   
## $air\_time  
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 20.0 82.0 129.0 150.7 192.0 695.0 9430   
##   
## $distance  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 17 502 872 1040 1389 4983   
##   
## $hour  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 9.00 13.00 13.18 17.00 23.00   
##   
## $minute  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 8.00 29.00 26.23 44.00 59.00   
##   
## $time\_hour  
## Min. 1st Qu. Median   
## "2013-01-01 05:00:00" "2013-04-04 13:00:00" "2013-07-03 10:00:00"   
## Mean 3rd Qu. Max.   
## "2013-07-03 05:22:54" "2013-10-01 07:00:00" "2013-12-31 23:00:00"

#6  
iris %>%  
 map(summary)

## $Sepal.Length  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.300 5.100 5.800 5.843 6.400 7.900   
##   
## $Sepal.Width  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 2.800 3.000 3.057 3.300 4.400   
##   
## $Petal.Length  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.600 4.350 3.758 5.100 6.900   
##   
## $Petal.Width  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.100 0.300 1.300 1.199 1.800 2.500   
##   
## $Species  
## setosa versicolor virginica   
## 50 50 50

QQ <- lm(Sepal.Length ~ Sepal.Width, iris)  
QQ

##   
## Call:  
## lm(formula = Sepal.Length ~ Sepal.Width, data = iris)  
##   
## Coefficients:  
## (Intercept) Sepal.Width   
## 6.5262 -0.2234

#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(V)

## [1] 7

#7b  
V[is.na(V)]

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

#7c  
V[3]

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

#7e  
str(V[7])

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